UNIVERSITY OF INSUBRIA
Department of Biotechnology and Life Sciences

Ph.D. School:
Biological and Medical Sciences

Ph.D. Program:
"Analysis, Protection and Management of Biodiversity"

- XXV Course -

PATTERNS OF POPULATION STRUCTURE AND ADAPTIVE GENETIC VARIATION IN ALPINE POPULATIONS OF *Picea abies* (L.) KARST.

By
Erica Adele Di Pierro

Advisor: Prof. Giorgio Binelli
Co-advisor: dott. Nicola La Porta

Ph.D. Dissertation

December, 2012

SUMMARY

Forest trees dominate many alpine landscapes and are currently exposed to changing climate. *Picea abies* (L.) Karst (Norway spruce) is one of the most important conifer species of the Italian Alps due to its ecological and economical relevance. Natural populations of this species are found across steep environmental gradients with large differences in temperature and moisture availability. Steep environmental gradients represent interesting models to study the interaction between natural selection and gene flow, especially when aiming to better understand adaptation processes under global change. The present work aims to understand adaptive responses to changing climate by determining and quantifying patterns of genetic diversity in natural population of *P. abies.* A wide array of potential candidate genes was tested, by means of Single Nucleotide Polymorphisms (SNPs), for correlation with climatic and environmental parameters at different spatial scales: i) a geographical scale corresponding to the natural distribution of *P. abies* across the Italian Alps and ii) at a regional scale on the Eastern Italian Alps. Weak population structure was revealed at the geographical scale with only one population clearly divergent from the unique major genetic cluster identified. At the regional scale, hierarchical analyses of molecular variance revealed that most of the genetic variability was found within populations (ca. 99%), and small but significant variation was also found due landscape features (ca. 0.38%). In order to detect potentially adaptive markers, classical F_{ST} outlier approaches were first applied and five outlier loci were revealed at broad scale, while contrasting results were obtained at the regional scale according to the model used. Subsequently, environmental association analyses were performed: at the geographical scale temperature and precipitation were found to influence allelic variation at seven polymorphic loci, while at the regional scale, the Alpine topography resulted a potential adaptive determinants at 19 polymorphic loci, thus considered of ecological relevance. The results obtained in this study may provide relevant information for forestry management and genetic conservation, to understand and quantify the effect of climate change on conifer species as well as their adaptive potential.

Contents

CHAPTER 1

Introduction

1.1 Forest Ecosystem

About one-third of the total land area in the world is covered by forests, including natural and artificial forests (Abril *et al.* 2011). Forest ecosystems play essential roles such as providing renewable raw materials, energy, maintaining biodiversity, and protecting land and water resources (Neale & Kremer 2011). Many forest trees, including gymnosperms such as cycads and conifers, are among the most ancient seed plants, first recorded as fossils in the Upper Devonian (350 million years ago) (Biswas & Jorhri, 1997). Conifers represent an extensive group with an important ecological role in terrestrial ecosystems, including some species with high commercial value, for example Norway spruce (*Picea abies*), Maritime pine (*Pinus pinaster*), Scots pine (*Pinus sylvestris*), Loblolly pine (*Pinus taeda*), Japanese cedar (*Cryptomeria japonica*) and Douglas fir (*Pseudotsuga menziesii*). Despite the extensive usage of forests by men in the last millennia, which could have caused genetic erosion, natural populations of trees revealed higher genetic variation than other plant species (Hamrick *et al.*, 1992; Nybom & Bartish, 2000, González-Martínez *et al.* 2006a). Conifer forests are characterized by large open-pollinated native populations and high levels of both genetic and phenotypic variation,

which together with their low domestication and ancient evolutionary history make them almost ideal models for the study of adaptive evolution using population genomics approaches.

1.2 *Picea abies* (L.) Karst.

1.2.1 Species ecology

Picea abies (L.) Karst is a conifer species belonging to the family Pinaceae, genus *Picea* and commonly known as Norway spruce. It is an ecologically and economically important forest tree species with a wide and heterogeneous distribution range. It is one of the most common tree species in Europe, where its natural range extends longitudinally from the French Alps to the Eastern Siberia and latitudinally from the Balkan Peninsula to Scandinavia (Schmidt-Vogt 1974). As demonstrated by several studies conducted by means of genetic markers and/or fossil records (Bucci & Vendramin 2000; Latałowa & Van der Knaap 2006; Heuertz *et al.* 2006; Tollefsrud *et al.* 2008), its current distribution is formed by three main domains derived from the post-glacial recolonization: the Alpine, the Hercyno-Carpathian and the Baltic-Nordic domain (Fig. 1.1).

P.abies is a significant species in various ecosystems reaching from the lowlands up to the subalpine vegetation zones. It is a continental forest species, which tolerates high summer temperatures but starts bud and shoot growth at relatively low temperatures (Partanen *et al.* 1998). It prefers moist soils with high seasonal water supply (Sutinen *et al.* 2002). Natural populations of this species may occur in pure stands, transitional stands mixed with Scots pine (*Pinus Sylvestris*), or mixed stands with European beech (*Fagus sylvatica*) and European silver fir (*Abies alba*). Norway spruce is wind pollinated and

seeds, which ripen in late autumn the same year, are also wind dispersed. The life span of Norway spruce is ca. 200 years, but at the northern limits of its range it may reach ca. 400 years.

Picea abies forests have been widely managed by men and extensively sown in Central Europe, particularly during the last two centuries, therefore large areas with artificial forests exist outside the natural habitat (Schmidt-Vogt 1986). Norway spruce wood is strong, soft, straight and fine-grained, and easily worked (Safford 1974) and indeed accounts for 80% of timber consumed in Europe.

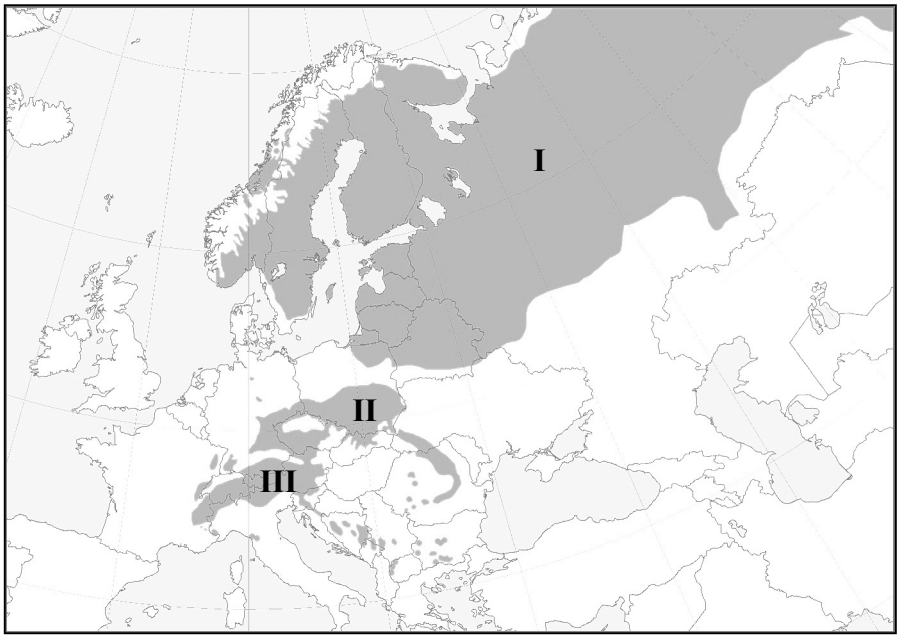

Fig.1.1 Present geographical distribution of *Picea abies* (grey pattern). Roman numbers mark the three main domains derived from the post-glacial recolonization: I) Baltic-Nordic domain, II) Hercyno-Carpathian domain, and III) Alpine domain (image adapted from EUFORGEN 2009, www.euforgen.org).

The Alpine domain of Norway spruce distribution is represented mainly by mountainous and subalpine locations such as the Alps, the Alpine foreland, the French and Swiss Jura, the Black forest, Bavarian forest and several mountain ranges of the Balkan

Peninsula including the Rhodopes. The Italian distribution of natural stands of Norway spruce ranges across the entire Alps, and includes a single stand located in the northern Apennines, near Campolino (Borghetti *et al.* 1988).

1.2.2 Genetic characterisation of Picea abies: a model for coniferous species

Conifer species are characterized by large size genomes and relatively low chromosome numbers (Murray 1998). The high presence of repetitive DNA is also a characteristic of most conifers (Murray *et al.* 2002). However, plants with large genome were generally found to possess a much higher proportion of repeated DNA sequences compared to those with smaller genome (Dean & Schmidt 1995). Chromosome number in Norway spruce is 24 (2n) (Khoshoo 1961) and the size of nuclear genome is about 4 x 10^{10} bp (Govindaraju & Cullis 1991). It is characterized by a high degree of complexity, with a significant portion of the genome being constituted of non-coding DNA regions (Schmidt *et al.* 2000) and also moderately to highly repetitive. Several classes of repeated sequences are observed within nuclear genome of *Picea abies*. Chloroplast genome represents one master chromosome of about 120 kb, distinctive of most conifers (Strauss *et al.* 1988; Raubeson & Jansen, 1992). Conifers cpDNA contains dispersed repetitive DNA that is associated with structural rearrangements (Hipkins *et al.* 1994). The genes required for dark synthesis of chlorophyll (a peculiar function of conifers and absent in other land plants) are present in the cpDNA of *Picea abies* (Lidholm & Gustafsson 1991). The gene content, gene structure, haplotype variation and repeat structure of the Norway spruce genome are currently under extensive study, as a recent and ambitious project has recently started with the aim of sequencing and assembling the Norway spruce entire genome: "The Spruce Genome Project" (http://www.congenie.org/).

Studies on the genetic structure of coniferous species, using different types of neutral markers, have shown low levels of population differentiation, even at a large scale. For example, allozyme studies of coniferous trees have revealed high levels of genetic variability and little population differentiation, as indicated by estimates of F_{ST} = 0.05 – 0.06 across the whole European range (Ledig 1986; Lagercrantz & Ryman 1990; Hamrick *et al.* 1992; Müller-Starck *et al.* 1992). This observed pattern of genetic structure in conifers has been attributed to the large population size, the essentially outcrossing system and the potential for long-distance gene flow (Hamrick *et al.* 1992; Ledig 1986). These characteristics foster the maintenance of genetic variability and act as cohesive factors preventing pronounced differentiation among sub-populations (Hamrick *et al.* 1992; Ledig 1986; Lagercrantz & Ryman 1990). Like other conifers, Norway spruce exhibits a relatively large amount of genetic variability and little differentiation among populations, however Lagercrantz and Ryman (1990) suggested that Norway spruce populations show a clear geographical differentiation largely reflecting relatively recent historical events related to the last glaciation and that this species is still in a process of adaptation and differentiation. The adaptive potential of a population and its ability to survive in a changing environment is largely due to its genetic variation. Several studies (Gömöry *et al.* 2011; Chen *et al.* 2012; Acheré *et al.* 2005) have already demonstrated Norway spruce ability to well adapt to different environmental gradients, and its natural variation has been observed at morphological, and physiological adaptive traits as well as with molecular markers (Müller-Starck *et al.* 1992; Krutovsky & Bergmann 1995). This adaptive potential and its distribution in such diverse habitats make it a very interesting species for studying the adaptive genetic variation of conifer forests in natural environment.

1.3 Climate change and effects on forests

The world's climate is changing. During the past 100 years, average global temperature has risen by about 0.74°C and is projected to increase from 1.8°C to 4°C until the next century (IPCC 2007). Although global climate change does occur naturally, the rate and magnitude of changes been seeing nowadays are believed to be both faster and larger than has ever occurred before. Human activities, such as burning of fossil fuels, deforestation and animal production, which increases the concentration of greenhouse gases in the atmosphere, are likely to be one of the main causes of this unprecedented change in the Earth's climate. These gases, such as carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O), may remain in the atmosphere from a decade to centuries; they act to heat the planet because of absorption and re-radiation of infrared radiation (IR). Even if human-caused greenhouse gas concentrations were to be stabilized, warming due to greenhouse gases and aerosols would continue for centuries due to the time scales associated with climate processes and feedbacks (IPCC 2007).

Climate change over the past ~30 years has produced numerous shifts in the distribution and abundances of the species (Root *et al.* 2003). Primary concern is nowadays arising about the high rate of climatic changes (Parmesan 2006; IPCC 2007) and considerable attention has been given to the possible effects on trees and forests. Studies of range shifts over the last 25,000 years showed that species ranges have migrated in close correlation with global climatic cycles. Conifer species, particularly in mountainous areas such as the Alps, may be able to face modest rates of climate change, by migrating short distances among microsites, slope aspects, or elevations. However, the predicted high rates of climatic change are likely to exceed the scale of local environmental heterogeneity in the longer term (Aitken *et al.* 2008). According to Reusch and Wood (2007) three are the

ways in which populations can react to rapidly changing environment: be plastic, migrate, or evolve and adapt; otherwise they will go extinct. Adaptive responses may affect species persistence, migration rate, and forest productivity, for this reason more complete understanding of adaptive responses to climate must be pursed (Davis *et al.* 2005)

In the last decade common garden experiments have provided strong evidences of the capacity of populations to locally adapt to climate (Howe *et al.* 2003; Savolainen *et al.* 2007; Aitken *et al.* 2008). However, adaptive responses to climate, while clearly important are not yet well understood (Davis *et al.* 2005).

Ongoing climate change has increased interest in the ability of species and populations to adapt to new environmental conditions (IPCC 2007). Previous studies suggest shifts in species distributions and migration patterns (Berthold *et al.* 1992; Bradshaw *et al.* 2004; Hari *et al.* 2006) as major consequences of climate change. Thus, monitoring of candidate gene allele frequencies along genetic clines may provide effective information on influence of climate change on natural populations (Narum *et al.* 2010).

1.4 Studying adaptation in natural forest-tree populations

The balance between gene flow and selection determines the extent of local adaptation; indeed genetic variation, dispersal and establishment rates may affect the potential of organisms to adapt to current climate change (Savolainen *et al.* 2007). The adaptive potential of non-model species, such as conifers, is very difficult to study especially for natural populations. This is due to several factors, first of all the long generation times characterizing conifer trees and uncontrolled mating. Moreover, the

inability to isolate variables in complex environments may produce confounding effects and the lack of genome information is also limiting (Narum *et al.* 2010).

1.4.1 Population genomics approaches

Most variation in adaptive traits is based on loci with small effects. Population genetics approaches and improved genomic resources may provide useful methods for the identification of loci responsible for this variation. Traditional methods such as provenance tests and screening of molecular genetic markers have been used to study and measure adaptive genetic diversity in forest-tree populations (Namkoong 2001). However, adaptive traits are usually under multigenic control. Developments in forest genomics provided new tools to identify the genes controlling adaptive traits. Population genetics approaches addressing evolution are combined with genome-wide sampling (Luikart *et al.* 2003; González-Martínez *et al.* 2006a) to study adaptive genetic variation in forest trees under new integrated population genomic approaches. Association mapping and genome scans represent recent valid tools of population analysis that together with recently developed functional markers (Andersen & Lubberstedt 2003) allow population genomics to reveal adaptive patterns in nature (González-Martínez *et al.* 2006a). Moreover, both transfer of information from model species (i.e., genes of known function in model systems) and standard neutrality tests, applied to population nucleotide sequence data of a single or a few genes, allow pre-selecting putative candidate gene loci for particular adaptive traits (González-Martínez *et al.* 2006a,b). In pines, the majority of genes that showed a departure from neutrality in DNA-sequence studies have been found related to biotic- and abiotic-

stress tolerance or key metabolic pathways (González-Martínez *et al.* 2006b; Pot *et al.* 2005; Mosca *et al.* 2012b).

1.4.2 Single Nucleotide Polymorphism (SNP) markers

Genetic markers are defined as heritable polymorphic characters that simply reflect differences in DNA sequences directly at the nucleotide level or indirectly at the level of gene expression. The ideal molecular marker for studying adaptive variation should be directly involved in the genetic control of adaptive traits; have an identified DNA sequence and known function; and have easily identifiable allelic variation. However, traditional markers hardly satisfy all these criteria. For this reason, new sequence-based markers are rapidly being developed (Andersen & Lubberstedt 2003) also in several forest-tree species (Brown *et al.* 2004). A new type of functional genomic marker is represented by Expressed Sequence Tags (EST): short (~200–700 nucleotides) cDNA sequences used to tag the gene from which the transcribed mRNA of a specific tissue originated (Bouck & Vision 2007). Single Nucleotide Polymorphisms (SNPs) can be broadly defined as any single base substitution/indel in the genome of an individual and are potentially the best type of genetic marker because of their abundance in the genome and their potential association with disease and adaptive traits. Moreover, most SNPs are biallelic, facilitating the development of automated high-throughput SNP-genotyping methods (Kwok 2001; Hirschhorn & Daly 2005). The utilization of SNPs as genetic markers has recently received much attention in human genetic studies of, for example, gene mapping (Wang *et al.* 1998; The International SNP Map Working Group 2001) and human evolution (Cargill *et al.* 1999; Hacia *et al.* 1999), due to their high frequency, also in coding DNA sequences, and the possibility for using highly automated analysis systems (reviewed in Landegren *et*

al. 1998). Similarly, in other species for which large-scale genome projects are underway, SNPs are being identified at a rapid rate (Lindblad-Toh *et al.* 2000; Marklund *et al.* 2000; Hoskins *et al.* 2001). Typical SNP discovery projects are based on direct sequencing of amplicons from a set of individuals (the discovery panel) covering the range of variation of a given species (González-Martínez *et al*. 2006b; Krutovsky & Neale 2005; Pot *et al.* 2005; Mosca *et al.* 2012b). SNP variation in ESTs has been discovered for several forest species by implementing *in silico* SNP discovery in forest-tree EST databases available (e. g. loblolly pine, http://dendrome.ucdavis.edu/NealeLab/adept2/overview.php/; maritime pine, http://www.pierroton.inra.fr/genetics/Pinesnps). However, ascertainment bias is possible *in silico* SNP discovery due to the typically small number of individuals from a limited number of populations used to generate EST libraries (Brumfield *et al.* 2003; Luikart *et al.* 2003; Nielsen & Signorovitch 2003).

In addition to recent advances in DNA sequencing and microarray technologies (Meldrum 2000a,b), one of the key factors facilitating the rapid characterization of SNPs in well-studied species has been the public availability of large amounts of overlapping sequence data, thus enabling the identification of sequence polymorphisms with a reduced amount of laboratory work: so-called 'data-mining' (e.g. Buetow *et al.* 1999). With increasing genomic information available for non-model organisms, single-nucleotide polymorphisms have begun to see increased use as genetic markers for population genetic studies (e.g., Luikart *et al.* 2003; Morin *et al.* 2004). Thus, SNPs present several favourable attributes: they are the most abundant class of polymorphisms in genomes (Reich *et al.* 2001) they can be genotyped in automated systems more easily than can microsatellites, and their mutation dynamics are more easily modelled and less variable among loci. Finally, in a large suite of unlinked SNPs distributed across the genome, both functionally

neutral and adaptive markers may be effectively used within a single study. This combination of information provides a powerful approach to study questions in ecological genetics because both demographic processes (i.e., gene flow and genetic drift) and local adaptation (i.e., selection) may be inferred (Narum *et al.* 2010).

1.4.3 Outlier-detection approaches

Loci with unusually high or low levels of variation and differentiation are referred to as outlier loci. A powerful method to identify signature of selection in the genome is based on the detection of these putative selected loci (Luikart *et al.* 2003). One simple method is based on the comparison of differentiation estimates (F_{ST}, Wright 1969) for putatively neutral molecular markers (e.g SSRs) and candidate gene markers, such as SNPs or EST-based markers. Markers that show higher (or lower) differentiation than putatively neutral ones can be considered as being under diversifying (or stabilizing) selection (González-Martínez *et al.* 2006a). Several methods have been developed in recent years, which do not require screening of neutral molecular markers. These are based on the methodology introduced by Lewontin & Krakauer (1973), which uses the coalescent theory to build, by means of simulation, a neutral expectation of genetic divergence among populations. Subsequently, the methodology was further developed by several authors (Beaumont & Nichols 1996; Vitalis *et al.* 2001; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier *et al.* 2009; Bonhomme *et al.* 2010) and implemented in computer tools that allowed the use of both dominant or codominant markers. Among the most widely used are: FDIST / DFDIST (Beaumont & Nichols 1996), DETSEL (Vitalis *et al.* 2001, 2003), and BAYESCAN (Foll & Gaggiotti 2008), the latter being a Bayesian method developed as an extension of that proposed by Beaumont & Balding (2004) to estimate directly the

posterior probability that each locus is subject to selection. These methods have found application in many studies, even on non-model species (e.g., Bonin *et al.* 2006; Namroud *et al.* 2008; Pariset *et al.* 2009; Eckert *et al.* 2010; Prunier *et al.* 2011; Chen *et al.* 2012; Holliday *et al.* 2012; see also references in Luikart *et al.* 2003), and comparisons between the different methods have also been made (Pérez-Figueroa *et al.*, 2010; Narum & Hess 2011; Vilas *et al.* 2012). These studies suggested caution when interpreting the results obtained with outlier detection methods and advocated control for Type error I or multitest correction. Moreover, it may be extremely difficult to assess whether all variation in the genes that behave as outliers are truly under adaptive selection (Luikart *et al.* 2003). To overcome this drawback, these methods should be used in combination with association or correlation analysis, where the association between a marker and a trait or an environmental variable may often provide the best evidence for adaptive significance (Luikart *et al.* 2003). In addition association mapping and gene-expression studies, represent valid approaches to integrate outlier detection methods.

1.4.4 Environmental association analysis

A comprehensive understanding of the molecular basis of adaptation and of the evolutionary processes responsible for shaping gene diversity in forest trees requires the use of more complementary approaches and disciplines. Loci involved in local adaptation can potentially be identified by an unusual correlation between allele frequencies and important ecological variables, such as climatic variables or clines. The standard method to detect loci underlying adaptive responses to environmental factors is the environmental association analysis (Vasemägi & Primmer 2005; Holderegger *et al.* 2008; Coop *et al.* 2010; Eckert *et al.* 2010). However, geography may affect gene flow, as well as

environmental variation, causing confounding inferences of natural selection from correlations between environmental variables and allele frequencies. For example, populations that tend to be geographically proximate often share environmental variables (Novembre & Di Rienzo 2009); therefore neighbouring populations can rarely be treated as independent observations (Coop *et al.* 2010). This issue actually occurs also when using F_{ST} as a summary statistic to identify selected loci (Robertson 1975; Excoffier *et al.* 2009). A recently developed Bayesian approach (Hancock *et al.* 2008; Coop *et al.* 2010) allows performing an environmental association analysis that corrects for background levels of population structure and differences in sample size before searching for correlations between allele frequencies and environmental variables (Hancock *et al.* 2008; Coop *et al.* 2010).

The environmental association analysis, where genetic variation is correlated to environmental factors, widely characterizes landscape genetics approaches (Lowry 2010), which directly uses environmental data to detect molecular markers linked to or located within genomic regions under selection (Holderegger *et al.* 2008; Holderegger *et al.* 2010). The subject of landscape genetics is the understanding of how landscape features structure populations (Manel *et al.* 2003). Frequencies of alleles in native populations across heterogeneous environments are estimated and subsequently correlated with local environmental conditions as, for example, estimates of temperature, precipitation, slope, altitude or habitat type. Molecular markers (e.g., amplified fragment polymorphisms [AFLPs] or SNPs) significantly correlated with such variables are seen as linked to genomic regions influenced by these factors (Holderegger *et al.* 2008; Manel *et al.* 2010). Several statistical methods can be used for this purpose, such as simple linear regression or generalized linear models, however spatial genetic structure need to be accounted. So far,

the most common approach has been to apply Bayesian clustering to estimate the genetic structure of samples and then use landscape genetic analysis with logistic or linear regression within each genetic cluster separately (Eckert *et al.* 2010; Scalfi *et al. in preparation*). Recently, more sophisticated approaches to account for spatial genetic structure have been introduced (Poncet *et al.* 2010; Manel *et al.* 2010, 2012). New powerful landscape genomic approaches will be provided by advancement of genomic and geo-spatial technologies. The parallel use of genome scans with environmental data in the study of forest tree adaptation to climatic changes will provide distinct clues for selective forces acting on molecular markers of adaptive relevance in real landscapes (Holderegger *et al.* 2008).

1.5 Aim of the study

The current work aims to assess the adaptive potential of natural populations of forest trees in an evolutionary and ecological context such as the Alpine environment, highly susceptible to climate change. To investigate the effect of future climatic changes, evolutionary changes should also be taken into account. Thus, the present study takes advantage of the recent development in high-throughput genomic resources, and a population genomics approach on putative candidate genes was used. The association between environmental factors and genetic diversity in natural populations of *Picea abies* was inspected at two different spatial scales using Single Nucleotide Polymorphisms (SNPs) as genetic markers. Therefore, with particular concern for one of the most relevant species of the *Picea* genus the following questions were addressed: i) what are the patterns of population structure for the *Picea abies* species across the Italian Alps? ii) How does landscape features affect these patterns at a local scale? iii) What are the putative candidate

genes associated to the climatic and environmental parameters considered in the range of our populations? iv) Does adaptation occur along the altitudinal gradient?

The present work is part of a major study aiming to conduct genetic monitoring of the alpine ecosystems for the Trentino region: the ACE-SAP *(ALPINE ECOSYSTEMS IN A CHANGING ENVIRONMENT: BIODIVERSITY SENSITIVITY AND ADAPTIVE POTENTIAL)* project.

References

Abril N, Gion J-M, Kerner R *et al.* (2011) Proteomics research on forest trees, the most recalcitrant and orphan plant species. *Phytochemistry*, **72**, 1219–1242.

Acheré V, Favre JM, Besnard G, Jeandroz S (2005) Genomic organization of molecular differentiation in Norway spruce (Picea abies). *Molecular Ecology*, **14**, 3191–3201.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Andersen JR, Lübberstedt T (2003) Functional markers in plants. *Trends in plant science*, **8**, 554–560.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **263**, 1619–1626.

Berthold P, Helbig AJ, Mohr G, Querner U (1992) Rapid microevolution of migratory behaviour in a wild bird species. *Nature*, **360**, 668–670.

Biswas C, Johri BM (1997) *The gymnosperms*. Springer-Verlag.

Bonhomme M, Chevalet C, Servin B *et al.* (2010) Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics*, **186**, 241–262.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular biology and evolution*, **23**, 773–783.

Borghetti M, Giannini R, Menozzi P (1988) Geographic variation in cones of Norway spruce (*Picea abies* (L.) Karst.). *Silvae genetica*, **37**, 178–184.

Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular ecology*, **16**, 907–924.

Bradshaw WE, Zani PA, Holzapfel CM (2004) Adaptation to Temperate Climates. *Evolution*, **58**, 1748–1762.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15255–15260.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

Bucci G, Vendramin GG (2000) Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. *Molecular ecology*, **9**, 923–934.

Buetow KH, Edmonson MN, Cassidy AB (1999) Reliable identification of large numbers of candidate SNPs from public EST data. *Nature genetics*, **21**, 323–325.

Cargill M, Altshuler D, Ireland J *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature genetics*, **22**, 231–238.

Chen J, Källman T, Ma X *et al.* (2012) Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics*, **191**, 865–881.

Coop G, Witonsky D, Rienzo AD, Pritchard JK (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, **185**, 1411–1423.

Davis MB, Shaw RG, Etterson JR (2005) Evolutionary responses to changing climate. *Ecology*, **86**, 1704–1714.

Dean C, Schmidt R (1995) Plant Genomes: A Current Molecular Description. *Annual Review of Plant Physiology and Plant Molecular Biology*, **46**, 395–418.

Eckert AJ, Bower AD, González-Martínez SC *et al.* (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, *Pinaceae*). *Molecular ecology*, **19**, 3789–3805.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Gömöry D, Longauer R, Hlásny T *et al.* (2011) Adaptation to common optimum in different populations of Norway spruce (*Picea abies* Karst.). *European Journal of Forest Research*, **131**, 401–411.

González-Martínez SC, Krutovsky KV, Neale DB (2006 a) Forest-tree population genomics and adaptive evolution. *The New phytologist*, **170**, 227–238.

González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006 b) DNA Sequence Variation and Selection of Tag Single-Nucleotide Polymorphisms at Candidate Genes for Drought-Stress Response in *Pinus taeda* L. *Genetics*, **172**, 1915–1926.

Govindaraju DR, Cullis CA (1991) Modulation of genome size in plants: the influence of breeding systems and neighbourhood size. *Evolutionary Trends in Plants*, **5**, 43–51.

Hacia JG, Fan J-B, Ryder O *et al.* (1999) Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nature Genetics*, **22**, 164–167.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hancock AM, Witonsky DB, Gordon AS *et al.* (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS genetics*, **4**, e32.

Hari RE, Livingstone DM, Siber R, Burkhardt-Holm P, Güttinger H (2006) Consequences of climatic change for water temperature and brown trout populations in Alpine rivers and streams. *Global Change Biology*, **12**, 10–26.

Heuertz M, De Paoli E, Källman T *et al.* (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics*, **174**, 2095–2105.

Hipkins VD, Krutovskii KV, Strauss S (1994) Organelle genomes in conifers: structure, evolution and diversity. *International Journal of Forest Genetics*, **1**, 179–189.

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature reviews. Genetics*, **6**, 95–108.

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in Plant Science*, **15**, 675–683.

Holderegger R, Herrmann D, Poncet B *et al.* (2008) Land ahead: using genome scans to identify molecular markers of adaptive relevance. *Plant Ecology & Diversity*, **1**, 273–283.

Holliday JA, Wang T, Aitken S (2012) Predicting Adaptive Phenotypes From Multilocus Genotypes in Sitka Spruce (*Picea sitchensis*) Using Random Forest. *G3: Genes|Genomes|Genetics*, **2**, 1085–1093.

Hoskins RA, Phan AC, Naeemuddin M *et al.* (2001) Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. *Genome research*, **11**, 1100–1113.

Howe G, Aitken S, Neale D *et al.* (2003) From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany*, **81**, 1247–1266.

IPCC (Intergovernmental Panel on Climate Change) (2007) *Summary for policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller (eds.)]*. Cambridge University Press, Cambridge, UK, and New York.

Khoshoo TN (1961) Chromosome Numbers in Gymnosperms. *Silvae Genetica*, **10**, 1–7.

Krutovski KV, Bergmann F (1995) Introgressive hybridization and phylogenetic relationships between Norway*, Picea abies* (L.) Karst., and Siberian, P. *obovata* Ledeb., spruce species studied by isozyme loci. *Heredity*, **74**, 464–480.

Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics*, **171**, 2029–2041.

Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. *Annual review of genomics and human genetics*, **2**, 235–258.

Lagercrantz U, Ryman N (1990) Genetic Structure of Norway Spruce (*Picea abies*): Concordance of Morphological and Allozymic Variation. *Evolution*, **44**, 38.

Landegren U, Nilsson M, Kwok PY (1998) Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome research*, **8**, 769–776.

Latałowa M, Van der Knaap WO (2006) Late Quaternary expansion of Norway spruce *Picea abies* (L.) Karst. in Europe according to pollen data. *Quaternary Science Reviews*, **25**, 2780–2805.

Ledig FT (1986) Heterozygosity, heterosis, and fitness in outbreeding plants. In: *Conservation biology* , pp. 77–104. Sinauer, Sunderland, Mass.

Lewontin RC, Krakauer J (1973) Distribution of Gene Frequency as a Test of the Theory of the Selective Neutrality of Polymorphisms. *Genetics*, **74**, 175–195.

Lidholm J, Gustafsson P (1991) Homologues of the green algal gidA gene and the liverwort frxC gene are present on the chloroplast genomes of conifers. *Plant molecular biology*, **17**, 787–798.

Lindblad-Toh K, Winchester E, Daly MJ *et al.* (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genetics*, **24**, 381–386.

Lowry DB (2010) Landscape evolutionary genomics. *Biology Letters*, **6**, 502–504.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature reviews. Genetics*, **4**, 981–994.

Manel S, Gugerli F, Thuiller W *et al.* (2012) Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular ecology*, **21**, 3729–3738.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular ecology*, **19**, 3824–3835.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Marklund S, Tuggle CK, Rothschild MF (2000) Mapping of the CYP1A1, SSTR1 and TTF1 genes to pig chromosome 7q refines the porcine-human comparative map. *Animal genetics*, **31**, 318–321.

Meldrum D (2000 a) Automation for genomics, part one: preparation for sequencing. *Genome research*, **10**, 1081–1092.

Meldrum D (2000 b) Automation for Genomics, Part Two: Sequencers, Microarrays, and Future Trends. *Genome Research*, **10**, 1288–1303.

Morin PA, Luikart G, Wayne RK, The SNP workshop group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Mosca E, Eckert AJ, Di Pierro EA *et al.* (2012 a) The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular ecology*, **21**, 5530–5545.

Mosca E, Eckert AJ, Liechty JD *et al.* (2012 b) Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications*, **5**, 762–775.

Müller-Starck G, Baradat P, Bergmann F (1992) Genetic variation within European tree species. *New Forests*, **6**, 23–47.

Murray BG (1998) Nuclear DNA Amounts in Gymnosperms. *Annals of Botany*, **82**, 3–15.

Murray BG, Friesen N, Heslop-Harrison JSP (2002) Molecular cytogenetic analysis of *Podocarpus* and comparison with other gymnosperm species. *Annals of botany*, **89**, 483–489.

Namkoong G (2001) Forest genetics: pattern and complexity. *Canadian Journal of Forest Research*, **31**, 623–632.

Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, **17**, 3599–3613.

Narum SR, Campbell NR, Kozfkay CC, Meyer KA (2010) Adaptation of redband trout in desert and montane environments. *Molecular ecology*, **19**, 4622–4637.

Narum SR, Hess JE (2011) Comparison of F(ST) outlier tests for SNP loci under selection. *Molecular ecology resources*, **11 Suppl 1**, 184–194.

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nature reviews. Genetics*, **12**, 111–122.

Nielsen R, Signorovitch J (2003) Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical population biology*, **63**, 245–255.

Novembre J, Di Rienzo A (2009) Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics*, **10**, 745–755.

Nybom H, Bartish IV (2000) Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology, Evolution and Systematics*, **3**, 93–114.

Pariset L, Joost S, Marsan P, Valentini A, Econogene Consortium (EC) (2009) Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics*, **10**, 7.

Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–669.

Partanen J, Koski V, Hänninen H (1998) Effects of photoperiod and temperature on the timing of bud burst in Norway spruce (*Picea abies*). *Tree Physiology* **18**, 811–816.

Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular ecology*, **19**, 2896–2907.

Pot D, McMillan L, Echt C *et al.* (2005) Nucleotide variation in genes involved in wood formation in two pine species. *New Phytologist*, **167**, 101–112.

Prunier J, Laroche J, Beaulieu J, Bousquet J (2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular ecology*, **20**, 1702–1716.

Raubeson LA, Jansen RK (1992) A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochemical Systematics and Ecology*, **20**, 17–24.

Reich DE, Cargill M, Bolk S *et al.* (2001) Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.

Reusch TBH, Wood TE (2007) Molecular ecology of global change. *Molecular ecology*, **16**, 3973–3992.

Robertson A (1975) Remarks on the Lewontin-Krakauer test. *Genetics*, **80**, 396–396.

Root TL, Price JT, Hall KR *et al.* (2003) Fingerprints of global warming on wild animals and plants. *Nature*, **421**, 57–60.

Safford LO (1974) Picea, spruce. In: *Seeds of woody plants of the United States (Schopmeyer CS, tech. Coord)* . Forest Service: 587B597, Washington.

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 595–619.

Schmidt A, Doudrick RL, Heslop-Harrison JS, Schmidt T (2000) The contribution of short repeats of low sequence complexity to large conifer genomes. *Theoretical and Applied Genetics*, **101**, 7–14.

Schmidt-Vogt H (1974) Das natürliche Verbreitungsgebiet der Fichte (*Picea abies* [L.] Karst) in Eurasien. *Allgemeine Forst- und Jagdzeitung*, 145:185–197.

Schmidt-Vogt H (1986) *Die Fichte: ein Handbuch in zwei Bänden. Wachstum, Züchtung, Boden, Umwelt, Holz*. Parey.

Strauss SH, Palmer JD, Howe GT, Doerksen AH (1988) Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 3898–3902.

Sutinen R, Teirilä A, Pänttäjä M, Sutinen M-L (2002) Distribution and diversity of tree species with respect to soil electrical characteristics in Finnish Lapland. *Canadian Journal of Forest Research*, **32**, 1158–1170.

The International SNP Map Working Group (2001) *Nature*, **409**, 928–933.

Tollefsrud MM, Kissling R, Gugerli F *et al.* (2008) Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. *Molecular ecology*, **17**, 4134–4150.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular ecology*, **14**, 3623–3642.

Vilas A, Pérez-Figueroa A, Caballero A (2012) A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of evolutionary biology*, **25**, 1364–1376.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of Variation Across Marker Loci as Evidence of Selection. *Genetics*, **158**, 1811–1823.

Vitalis R, Dawson K, Boursot P, Belkhir K (2003) DetSel 1.0: A Computer Program to Detect Markers Responding to Selection. *Journal of Heredity*, **94**, 429–431.

Wang DG, Fan JB, Siao CJ *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science (New York, N.Y.)*, **280**, 1077–1082.

Wright S (1969) *Evolution and the Genetics of Populations, Volume 2: Theory of Gene Frequencies*. University of Chicago Press.

CHAPTER 2

Patterns of population structure and adaptive variation to climate across the Italian range of Norway spruce (*Picea abies* [L.] Karst)

Abstract

Forest trees dominate many alpine landscapes and are currently exposed to changing climate. *Picea abies* (L.) Karst (Norway spruce) is one of the most important conifer species of the Italian Alps due to its ecological and economical relevance. Natural populations of this species are found across steep environmental gradients with large differences in temperature and moisture availability. This study seeks to determine and quantify patterns of genetic diversity for this species toward understanding adaptive responses to changing climate. A wide array of potential candidate genes was tested for correlation with climatic parameters characterizing sampled populations. Needles were sampled from 24 natural stands across the Italian species range. Sample locations were recorded by GPS technology to retrieve climatic data for each site. Samples were genotyped for 384 selected single nucleotide polymorphisms (SNPs) from 285 genes. To avoid false-positive association between genotype and climate, population structure was investigated. Low levels of differentiation among populations were revealed by both F_{ST}-*multilocus* and pairwise F_{ST} analyses. Both the Bayesian clustering method and the

multivariate analysis validated this result, suggesting the presence of only one population clearly divergent from the unique major genetic cluster identified. The detection of selection on individual SNPs revealed five F_{ST} outliers, while an environmental association analysis detected seven SNPs associated to one or more climatic variables. Precipitation, more than temperature, was more often associated with genotype, thus reflecting the importance of water availability for Norway spruce. These findings could provide relevant information for forestry management and genetic conservation, to understand and quantify the effect of climate change on this species as well as its ability to genetically adapt.

Key words: alpine environment, environmental association analysis, *Picea abies*, population structure, SNPs.

2.1 Introduction

Climate is a major determinant of the composition and distribution of biomes due to species-specific physiological thresholds of temperature and precipitation tolerance (Walther *et al.* 2002). Climate acts as a potent selective force in natural populations, especially for those species unlikely to migrate fast enough to track rapidly changing climate, such as forest trees, that may need to rapidly adapt in place. High mountain regions such as the European Alps are particularly vulnerable to climatic changes (Theurillat & Guisan 2001); especially since the early 1980s the Alps have experienced above-average temperature increases (Beninston *et al.* 1997; Frei *et al.* 2010). Forest ecosystems form the dominant landscape in many alpine environments. The recent increase in global temperatures (IPCC 2007) has led to shifts in species distributions to higher altitudes and/or latitudes (Walther *et al.* 2002) to track environmental conditions and

prevent widespread extinction. Range expansions have been observed in a wide range of taxa (Walther *et al.* 2002; Hickling *et al.* 2006) with considerable evidence for these range shifts to occur in plant species (Jump & Peñuelas 2005; Buckley *et al.* 2012). Despite large amounts of gene flow and slow rates of evolution, rapid local adaptation to heterogeneous environments observed in vascular plants has long interested evolutionary biologists (Petit & Hampe 2006). Such adaptive processes result from selective pressures exerted by environmental changes and often reveal signatures of selection at the molecular level. Identifying associations between the causative environmental factors and the genotype is a challenging task, especially in non-model species (Luikart *et al*. 2003; Mariac *et al.* 2011; Le Corre & Kremer 2012). Despite recent advances in the development of genomic resources for several plant species (Arabidopsis Genome Initiative 2000; Jaillon *et al.* 2007; International Rice Genome Sequencing Project 2005; Tuskan *et al.* 2006) deciphering the molecular basis of plant adaptation is still a central challenge (González-Martínez *et al.* 2006; Storz & Wheat 2010; Neale & Kremer 2011). Genome scans for association with environmental variation is a powerful approach toward detecting genes and regions of the genome underlying adaptation, especially if accompanied by genotype to phenotype association studies for complex adaptive traits (Mariac *et al.* 2011).

Genetic variation can be searched for signatures of selection (Nielsen 2005; Storz 2005) in populations exposed to environmental differences or along environmental clines (Endler 1986). Such an approach depends on the availability of well-characterized genomes (Luikart *et al.* 2003); therefore it may be limited for non-model species where often random anonymous markers have been used (Murray & Hare 2006; Poncet *et al.* 2010; Mariac *et al.* 2011). Alternatively, single nucleotide polymorphisms (SNPs) from both coding and non-coding genomic regions allow for both historical demography

analysis and searches for natural selection. The recent development of genotyping methods for large numbers of individuals and many SNPs enables population genomic approaches for identifying loci that might be targets for selection, even in non-model organisms.

Norway spruce (*Picea abies* [L.] Karst) is an ecologically and economically important forest tree species with a broad and heterogeneous distribution in Europe. Its natural range extends longitudinally from the French Alps to the Eastern Siberia and latitudinally from the Balkan Peninsula to Scandinavia (Schmidt-Vogt 1974). Current natural distribution of Norway spruce has been strongly influenced by glacial periods. Studies conducted by means of genetic markers and/or fossil records revealed three main domains (Bucci & Vendramin 2000; Latałowa & van der Knaap 2006; Heuertz *et al.* 2006; Tollefsrud *et al.* 2008) derived from the recolonization routes after the last glaciations (Schmidt-Vogt 1977; Huntley & Birks 1983; Lagercrantz & Ryman 1990): the Baltic-Nordic domain, the Hercyno-Carpathian and the Alpine domain. In Italy, natural stands are distributed across the Alpine range with a single stand in the Northern Apennines, near Campolino. This marginal population has been suggested as the final step of the recolonization routes proposed by Lagercrantz and Ryman (1990) or, on the other hand, as a repopulation centre for the western Alps (allozymic data: Giannini *et al.* 1991; morphological data: Borghetti *et al.* 1988). It is a continental tree, which tolerates high summer temperatures but starts bud and shoot growth at relatively low temperatures (Partanen *et al.* 1998), and it prefers moist soils with high seasonal water supply (Sutinen *et al.* 2002). Studies on the genetic structure of Norway spruce populations, using different types of neutral markers, have shown, as for most coniferous species, low levels of population differentiation, but high within-population genetic variability (Lagercrantz & Ryman 1990; Bergmann & Ruetz 1991). The adaptive potential of a population and its

ability to survive in a changing environment is largely due to its genetic variation. Several studies (Gömöry *et al.* 2011; Chen *et al.* 2012; Acheré *et al.* 2005) have already demonstrated Norway spruce ability to well adapt to different environmental gradients. This adaptive potential and its distribution in such diverse habitats make of it a very interesting species for studying the climate effects on adaptive genetic variation.

Understanding the process of adaptation and quantifying standing genetic variation is a crucial point given the predicted climatic changes (IPCC 2007). The amount and extent of evolutionary adaptation can be limited by low genetic diversity, gene flow, and costs associated with adaptive change. Sessile organisms, such as plants, are physically confined to their habitat, therefore forced to deal directly with climate change. The present study aims to inspect the genetic basis that allows these organisms to adapt and respond to climate change, and provide a new cohort of single nucleotide polymorphisms to be assayed in natural populations for conservation genetics purposes or possibly even in breeding programs.

The association between climatic factors and genetic diversity was investigated across 24 Norway spruce natural populations, distributed across species natural range on the Italian Alps and Apennines. The sampling design was meant to first, allow the investigation of the population structure on a regional scale, and second, by deriving climate indicators for each sampling area, to look for association between the climatic parameters and polymorphic genes. A combined approach of selection scan and association analysis was conducted. To overcome the lack of genomic data on this species, putative candidate genes that might possibly be involved in abiotic or biotic adaptation, were selected in conserved regions detected in a very well studied conifer species, *Pinus taeda*, and 384 SNPs were selected to i) assess genetic structure of populations, ii) identify loci

that might be target for selection, iii) identify the link between the functional variation at the genome level and the spatial climatic variation (Fig. 2.1). This study may represent one further step towards the understanding of the adaptive variation of natural populations of forest trees. Specifically one of the most relevant species of the *Picea* (spruce) genus was here addressed in an evolutionary and ecological relevant context such as the Alpine environment, highly susceptible to climate change.

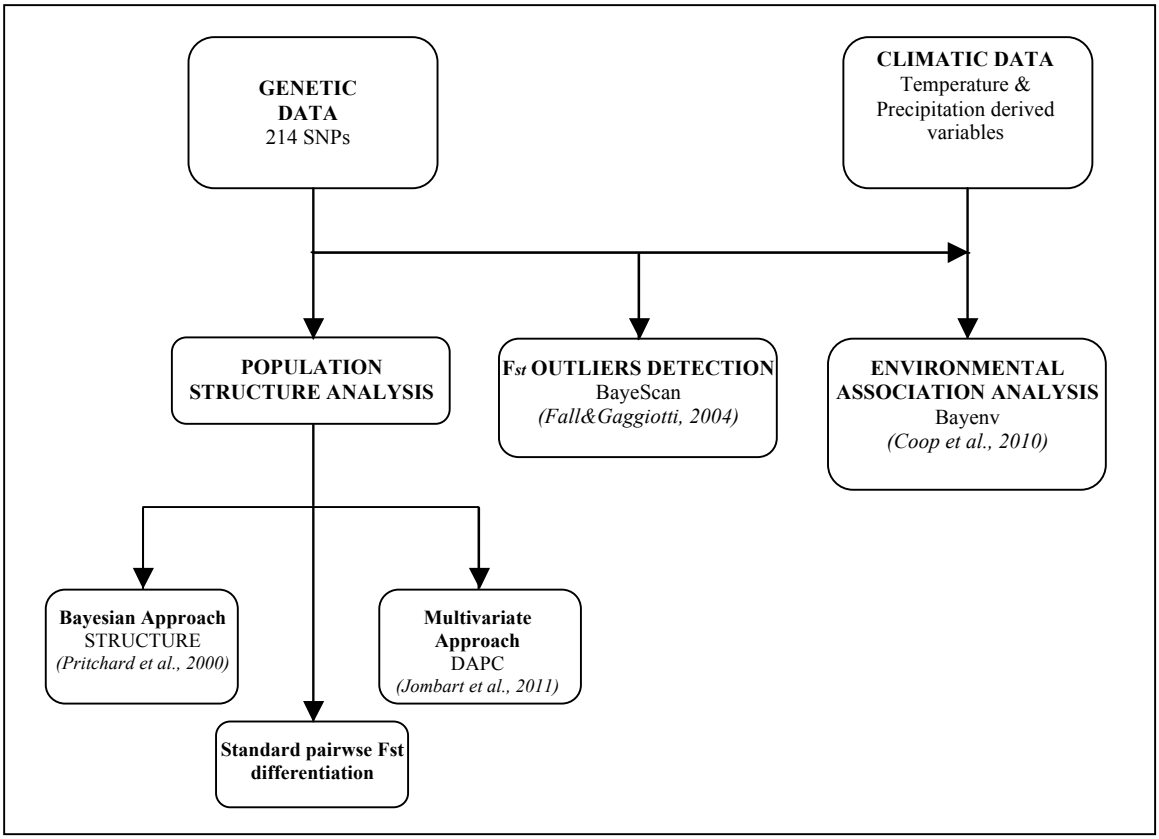

Fig. 2.1. The diagram summarises the subsequent steps (data generation, population structure analysis and adaptive variation identification) and combined approaches used across the study.

2.2 Material and methods

2.2.1 Plant material

A total of 860 individuals were sampled from 24 natural stands of Norway spruce (Fig. 2.2, Table 2.1). The sampling was conducted within the ACE-SAP (Alpine ecosystem in a Changing Environment: biodiversity Sensitivity and Adaptive Potential) project (see also Mosca *et al.* 2012), whose goal was studying biodiversity and adaptation in the region of Trentino. Thus sampling sites were mostly selected across the eastern Alps, with the exception of two populations: number 19 in the Maritime Alps (Valdieri) and number 24 (Campolino) in the northern Apennines. From 25 to 65 mature trees per population, at least 10 meters from each other, were sampled and the GPS position of each tree was recorded. A Trimble GPS device (www.trimble.com; Trimble®TerraSync™ software v3.20) was used and subsequently post-processed differential correction was applied to the data (GPS Pathfinder Office® software v4.10). The differential correction process improves the accuracy of the data by comparing the GPS position collected in the field with base data collected at a known location at the same time that the field data were collected.

Fresh needles of each tree were sampled and stored in labelled snap-vials immediately after collection along with two 0.5g silica gel desiccant packets. Tissue was then dried at 64 °C degrees and stored at room temperature until DNA extraction. Desiccated needle tissue (50 mg) was cut into 1 cm segments and placed in a 2 ml Deep Well Plate (DWP). DWP containing samples and grinding balls were then frozen using liquid nitrogen and placed at -80 °C overnight. Five or less 32” stainless steel grinding balls were added to each well, and plates were placed in tissue homogenizer (Spex SamplePrep, Metuchen, NJ, USA) and ground for 30 second intervals at 950 strokes/min three times. DNA extraction was carried out by programming the standard DNeasy Plant

96 Kit (Qiagen) protocol on a Liquid Handling machine (Eppendorf). DNA was then quantified using Quant-iT PicoGreen Assay Kit (Invitrogen) on a fluorescent plate reader (PerkinElmer).

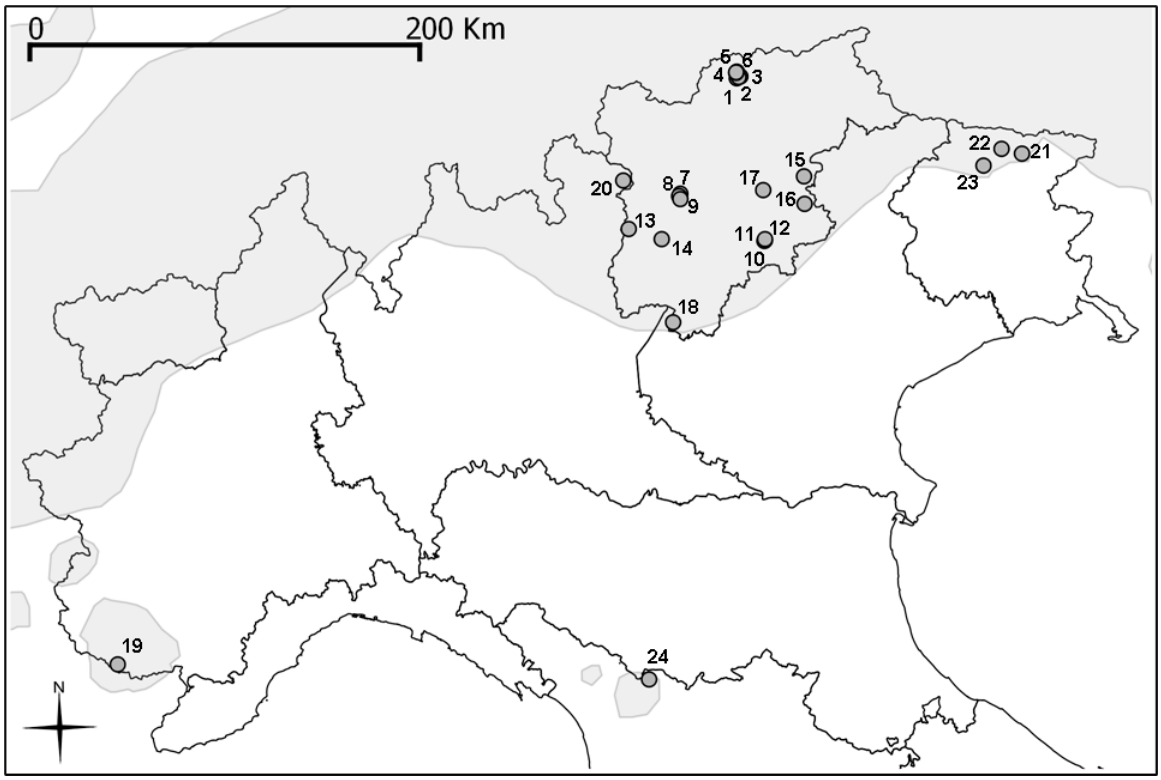

Fig 2.2. Sampled populations across the natural range of Norway spruce (grey shadow, modified from EUFORGEN network distribution maps) on the Italian Alps. The map was realized using Qgis 7.1 software.

Table 2.1. List of sampled populations with their geographical location and elevation.

Population	n†	Long	Lat	Locality	Region*	Altitude (m)
1	25	11.360	46.887	Val Ridanna	AA	1109
2	25	11.348	46.883	Val Ridanna	AA	1499
3	25	11.337	46.879	Val Ridanna	AA	1692
4	65	11.332	46.903	Val Ridanna	AA	1242
5	25	11.336	46.906	Val Ridanna	AA	1489
6	65	11.339	46.909	Val Ridanna	AA	1701
7	25	10.938	46.354	Val di Sole	TR	775
8	25	10.931	46.344	Val di Sole	TR	1084
9	25	10.940	46.330	Val di Sole	TR	1517
10	35	11.488	46.125	Val Calamento	TR	1184
11	25	11.491	46.127	Val Calamento	TR	1306
12	35	11.492	46.132	Val Calamento	TR	1455
13	65	10.591	46.195	Val Genova	TR	1683
14	65	10.809	46.146	Val d'Algone	TR	1757
15	65	11.766	46.418	Val San Nicolo'	TR	1879
16	65	11.764	46.291	Paneveggio	TR	1805
17	25	11.492	46.359	Passo Lavaze'	TR	1799
18	25	10.874	45.761	Avio	TR	1544
19	25	7.268	44.182	Valdieri (CN)	PM	1598
20	25	10.564	46.421	Val Forni (SO)	LO	2219
21	25	13.218	46.476	Vault-Moggio (UD)	FVG	1743
22	25	13.087	46.504	Salino-Paularo (UD)	FVG	1554
23	25	12.958	46.431	Villa Santina (UD)	FVG	1003
24	25	10.665	44.115	Campolino (PT)	TS	1575

†n: number of individuals sampled in each population
*AA: Alto-Adige, TR: Trentino, PM: Piemonte, LO: Lombardia, FVG: Friuli-Venezia Giulia, TS: Toscana.

2.2.2. SNPs selection and genotyping

Re-sequencing and SNP discovery was conducted as part of the Comparative Re-Sequencing in Pinaceae project (CRSP; http://dendrome.ucdavis.edu/NealeLab/crsp/). A sub-set of 1,024 primer-pairs derived from loblolly pine EST unigenes (ADEPT2 project http://dendrome.ucdavis.edu/NealeLab/adept2/overview.php/) was used to re-sequence (Sanger & Coulson 1975) haploid megagametophyte DNA samples from 12 Norway spruce trees across their natural range (Table S1 - Appendix A). The re-sequencing data

were processed using PineSAP, a customized pipeline performing sequence alignment and SNP identification (Wegrzyn *et al.* 2009), and resulted in the discovery of 3,417 SNPs (CRSP, http://dendrome.ucdavis.edu/NealeLab/crsp/overview).

In the present study a subset of 384 SNPs entirely derived from the CRPS re-sequencing panel was used. These SNPs, belonging to 285 putative candidate genes that might be involved in abiotic or biotic adaptation, were selected for genotyping based mainly on five factors: (i) Illumina design score, (ii) minor allele frequency in the re-sequencing panel, (iii) visual checking of the re-sequencing data alignments, (iv) SNP annotation and (v) maximization of gene coverage, hence from one to a maximum of four SNPs per gene were selected (Table S2 – Appendix A). Genotyping was carried out at the DNA Technologies Core in the UC Davis Genome Center, Davis CA, USA (http://www.genomecenter.ucdavis.edu) using the Illumina BeadXpress platform which supports the GoldenGate Genotyping assay (Shen *et al.* 2005) on digitally inscribed micro-beads *VeraCode (*Lin *et al.* 2009*)*. Intensity data for each SNP were then quantified and matched to specific alleles using GenomeStudio V2009.1 (Illumina). For SNPs to be included in the final data set, conservative thresholds of 0.25 and 0.80 for the $GenCall_{50}$ (GC_{50}) and call rate (*CR*) indexes, respectively, were used. These are quality scores to assess the validity of Illumina genotyping data (Shen *et al.* 2005*;* Eckert *et al.* 2009) and describe the accuracy of samples to be grouped into genotypic clusters (GC_{50}) and the portion of total sample that was genotyped for a given SNP (*CR*). Another measure of the reliability of SNP detection is the *GenTrain* score, which is based on the distribution of genotyping classes (Pavy *et al.* 2008). SNPs retained had a minimum *GenTrain* score of 0.50. All data were visually checked and manually adjusted when errors in genotypic clusters were evident. For each SNP the minor allele frequency (*MAF*) and the Wright's

inbreeding coefficient ($F = 1 - H_O/H_E$) were calculated. Only polymorphic loci were included in the calculations. Quality control steps were performed removing loci with absolute values of $F > 0.30$ and/or $MAF < 0.1$ from the final data set. Lastly, to evaluate the level of linkage disequilibrium (LD), allelic correlation (r^2) was estimated. Analyses were conducted using the GENETICS package in R (R Development Core Team 2010).

2.2.3 Genetic diversity and population structure

For each population, genetic diversity was estimated by calculating the following indices: mean number of alleles per SNP (A), observed (H_O) and expected (H_E) heterozygosity (Nei 1978) and the within-population fixation index (F_{IS}), its deviation from zero was tested by 10,000 permutations of allele frequencies within population. Furthermore, a global value of multilocus differentiation (F_{ST}-*multilocus*) was estimated, following standard ANOVA as in Weir & Cockerham (1984) with 95% confidence interval (CI) defined by 1,000 permutations. All calculations were performed using the GENETIX 4.05.2 software (Belkhir *et al.* 1996-2004).

Three statistical approaches were used to investigate population genetic structure (Fig. 2.1). First, pair-wise F_{ST} values between sampled populations were estimated using GENETIX 4.05.2 (Belkhir *et al.* 1996-2004). Then, the Bayesian clustering approach implemented in the program STRUCTURE v.2.3 (Pritchard *et al.* 2000; Hubisz *et al.* 2009) was used to infer individual membership to one or more genetic clusters on the basis of their multilocus genotype. One to 19 putative genetic clusters (k) were tested by executing 25 independent STRUCTURE runs for each k value. Each run was carried out with 1,000,000 MCMC steps following a burn-in period of 100,000 iterations. An admixture model with correlated allele frequencies was chosen and the degree of admixture alpha,

inferred from the data, was used. The parameter of allelic frequencies distribution lambda (λ) was set at the mean value of λ estimates inferred in three different runs with $k = 1$. The results from all runs were subsequently summarized and assessed using STRUCTURE HARVESTER v.0.6.92 (Earl & vonHoldt 2012). To determine the optimal number of inferred k, Δk (Evanno *et al.* 2005) was evaluated. Averaged admixture coefficients were then estimated for each k value using the *Greedy* algorithm with 1,000 random input orders as implemented in CLUMPP (Jakobsson & Rosenberg 2007). The clustering results were visualized using bar plots with the software DISTRUCT (Rosenberg 2004). This procedure was followed to run STRUCTURE on the same dataset twice: (i) by applying the basic model implemented in STRUCTURE v2.2 (Pritchard *et al.* 2000), which uses only the genetic information to identify the population structure; and (ii) by using prior population information through the LOCPRIOR model (Husbiz *et al.* 2009). This model was shown to better infer ancestry estimates when signal of structure is too weak, while giving similar results to those of the basic model when the dataset is highly informative (Husbiz *et al.* 2009; Li *et al.* 2010). Results obtained from the (ii) model are reported, which are biologically meaningful and easier to interpret, although do not present any substantial difference with the results from the (i) model.

To validate the results obtained by STRUCTURE, a Discriminant Analysis of Principal Components (DAPC: Jombart *et al.* 2010) was also performed by the *adegenet* 1.3-0 package (Jombart 2008) for the free and open source R software (R Developmental Core Team 2010). According to Jombart *et al.* (2010) this new multivariate method is free from assumptions regarding the population genetic models or the data structure, and may be more suitable in detecting hierarchical structure and clinal variation. To search for evidence of genetic clusters in DAPC, the sequential *k–means* clustering algorithm

(*find.clusters* function) was used, retaining all principal components. The Bayesian Information Criterion (BIC) was used to infer the optimal number of clusters: the lowest BIC value after which BIC increased or decreased by the least amount. To describe the identified clusters, DAPC executes a first step in which the genetic data are scaled and centred using principal components analysis, and a second step where the retained principal components are then submitted to a linear discriminant analysis, so that within-group variation is minimized and among-group variation is maximized. When applying DAPC to assign individuals into clusters, the number of principal components incorporating 80% of the cumulative variance was retained. Further, DAPC was also performed on the same data-set, using sampling sites as pre-defined groups and retaining the number of principal components encompassing about the 70% of the cumulative variance.

2.2.4 Climatic data

In order to characterize each sampling site for climatic indicators, climate data were obtained from two different data sources. Monthly and annual cumulative precipitation estimates for the period 1981-2010 were derived from the European Climate Assessment & Dataset (Haylock *et al.* 2008) time series, with a resolution of 0.25 decimal degree, while temperature data were obtained from daily reconstructed MODIS (Moderate Resolution Imaging Spectroradiometer) LST (Land Surface Temperature) data from the Terra and Aqua satellites (original data available at ftp://e4ftl01.cr.usgs.gov/). In particular, the reconstructed time series from Neteler (2005, 2010) were used in this study to derive monthly and annual minimum and maximum temperature for the period 2002-2011 at a resolution of 200 m pixel. Climatic variables and elevation were derived for the study

areas using GRASS GIS 6.4 (GRASS Development Team 2011, http://grass.osgeo.org; Neteler *et al.* 2012). Georeferenced positions of each sampling sites were recorded on the field. Sets of months using quarters, determined to the nearest month, were also defined, and classified as: coldest, warmest, driest and wettest quarter. In order to estimate climatic indicators across aligned time series, the last 10 years were used also for precipitation data, due to the restricted range (10 years) of high-resolution temperature data available. However, the 10 years of annual precipitation trend was first checked to appropriately match the 30 years trend using a two-sample Kolmogorov-Smirnov test. Subsequently, a Spearman's rank order correlation coefficient (r_s) was used to test for correlation between the climatic variables.

2.2.5 F_{ST} outlier analysis

The Bayesian approach for detecting outlier loci, implemented by the program BayeScan v2.1 (Foll & Gaggiotti 2008), was used to identify potentially selected loci. This method directly estimates the probability of a locus being under selection from the distribution of locus-specific F_{ST} and takes all loci into account. In this Bayesian approach, allele frequencies are assumed to follow a Dirichlet distribution, and the posterior distributions of both a model with selection and a model without selection are simultaneously estimated by using a reverse-jumping Markov Chain Monte-Carlo (MCMC) approach. Run parameters of 20 pilot runs of 500 length and 50,000 MCMC iterations with additional 50,000 burn-in steps were used. Outliers were defined at 5% and 1% significant levels of posterior probability corrected by False Discovery Rate (FDR) method (Benjamini & Hochberg 1995) as implemented in BayeScan. To account for population structure that considerably increases the number of false positives (Excoffier *et*

al. 2009), the same analysis was performed twice: first on the full dataset including all populations and second on a reduced dataset where population 19 was removed.

2.2.6 Environmental association analysis

To assess the correlation between SNP allele frequencies and climatic variables, a Bayesian generalized linear mixed model, recently developed by Coop *et al.* (2010) and implemented in the program Bayenv, was used. This method allows correcting for differences in sample sizes and background levels of population structure, responsible for confounding effects when searching for adaptive variation (Excoffier *et al.* 2009). To control for the correlation of allele frequencies across populations, a null model is estimated in a first step, providing a covariance matrix using a Monte Carlo Markov Chain (MCMC). The covariance matrix incorporates all random effects due to shared population history and gene flow, and informs the model as to how to weight the different populations when, on the second step, the effect of the environmental variables on the individual SNP frequency is tested (Hancock *et al.* 2008; Coop *et al.* 2010). In brief, this approach compared a model in which allele frequencies were dependent on a given environmental variable in addition to population structure, to a model in which allele frequencies were dependent on population structure only. Further details on the model and its application can be found in Hancock *et al.* (2010) and Eckert *et al.* (2010).

In the present study, the covariance matrix was first estimated on the entire dataset using all 214 SNPs. Multiple runs of the MCMC algorithm were carried out to check for convergence and matrices were compared within and across independent runs. In order to avoid any concern about the variation across the draws, the mean covariance matrix over convergent iteration was used for the test of correlation. As a measure of support for the

correlation between the allele frequency at each SNP and a specific climatic variable, a Bayesian factor was calculated as the ratio of the posterior probability under the environmental effect model to that under the null model. According to Kass & Raftery (1995), "substantial" evidence for selection is indicated by a BF of 3 while "strong" evidence is given by BF values larger than 10. Geography affects both gene flow and environmental variation causing further difficulties in inferring natural selection from the correlation between allele frequencies and environmental variables (Eckert *et al.* 2010), therefore the whole analysis was repeated for a reduced dataset where the most isolated and distant population, 19, was subsequently removed. Although such statistical approaches as the one implemented in Bayenv can be used to ameliorate the difficulties due to demographic history and geography effects, they might not always be effective. For this reason data where no population structure is evident, after removal of divergent populations, were further explored.

2.3 Results

2.3.1 Data summary and genotyping success

DNA genomic was successfully extracted from 836 trees of the 860 sampled (97%). Of the 384 SNPs genotyped using the Illumina BeadXpress platform, 288 (75%) SNPs were successfully genotyped according to the quality scores thresholds used (mean values of GC_{50}, *CR* and *GenTrain* being 0.65, 0.97 and 0.71 respectively) and 245 were polymorphic (Table 2.2). Estimates of *MAF* and *F* were calculated for each polymorphic SNP (Table S3 – Appendix A) and 31 SNPs failed to pass this quality control step (Table 2.2) and were thus excluded from further analyses. Therefore, the final dataset consisted of 214 SNPs, selected from 172 unique candidate genes and successfully genotyped for 826

individuals. The average rate of genotyping success for each population was over 95%, ranging from 80% in one population to 100% in 10 populations (Table 2.3).

Table 2.2. Genotyping success of SNPs markers and genes. The number of SNPs that failed to pass quality control (*MAF* < 0.1 and/or |*F*| > 0.30) is also shown.

Category	Number of SNPs	% SNPs	Number of genes	% genes
Failed	96	25	62	21.8
Successful MONOMORPHIC	43	11.2	28	9.8
Successful POLYMORPHIC	245	63.8	195	68.4
MAF < 0.1	18 + (1)†	5		
|*F*| > 0.30	12	3.1		
Grand Total	384	100	285	100

† One SNP was beyond both *MAF* and |*F*| thresholds

Estimates of LD in the final dataset were quite low with an average value of r^2 =0.003 and only eight pairs of SNPs out of ca. 23 x 10^3 pairwise comparisons having r^2 > 0.80. SNPs with high LD estimates were retained in further analyses.

2.3.2 Genetic diversity analysis and population structure

H_O and H_E estimates (Table 2.3) ranged from 0.237 to 0.275 (grand mean: 0.260 ± 0.009) and from 0.243 to 0.269 (grand mean: 0.260 ± 0.006), respectively. Within-population fixation index (F_{IS}) values were significantly different from zero in five populations (Table 2.3).

Overall, little genetic differentiation was detected with F_{ST}-*multilocus* of 0.012 (95% CI: 0.01042 - 0.01339). Pair-wise F_{ST} estimates between sampled populations ranged between 0.000 and 0.075, with the highest values involving the two disjoint populations, number 19 in the western Alps and number 24 on the Apennines (Table S4 – Appendix A).

In agreement with the overall genetic differentiation and the values of pair-wise F_{ST} found, the Bayesian clustering analysis revealed a weak population structure with a clear divergence of population 19 (Fig. 2.3).

Table 2.3. Genetic parameters estimated for the sampled populations. Mean number of alleles per SNP: A, average observed heterozygosity: H_O, average unbiased expected heterozygosity: H_E, within – populations fixation index: F_{IS}, standard deviation: SD. *$P \leq$ 0.05, ***$P \leq 0.001$ based on 10,000 permutation between individuals within populations.

Population	% of genotyped samples	A	H_O	H_E	F_{IS}
1	88	1.94	0.275	0.269	-0.024
2	100	1.96	0.260	0.264	0.016
3	100	1.93	0.266	0.265	-0.007
4	100	1.98	0.269	0.268	-0.006
5	100	1.91	0.263	0.258	-0.020
6	100	1.98	0.267	0.266	-0.005
7	100	1.94	0.267	0.259	-0.032*
8	100	1.93	0.251	0.253	0.009
9	100	1.91	0.260	0.258	-0.008
10	85.71	1.93	0.271	0.266	-0.020
11	96	1.93	0.256	0.259	0.013
12	88.57	1.95	0.267	0.263	-0.016
13	98.46	1.97	0.268	0.266	-0.007
14	95.38	1.99	0.257	0.256	-0.005
15	98.46	1.99	0.260	0.266	0.022 *
16	98.46	1.98	0.262	0.262	0.000
17	96	1.93	0.272	0.263	-0.036*
18	88	1.91	0.250	0.256	0.025
19	88	1.85	0.237	0.243	0.023
20	100	1.93	0.259	0.255	-0.018
21	96	1.92	0.248	0.266	0.069 ***
22	100	1.94	0.253	0.261	0.028*
23	80	1.91	0.262	0.261	-0.005
24	92	1.88	0.245	0.249	0.018
Average	95.38	1.94	0.260	0.260	0.001
SD	5.88	0.03	0.009	0.006	0.023

According to Δ*k* (Evanno *et al.* 2005) the presence of two main genetic pools of origin in our sampled individuals was inferred by STRUCTURE v. 2.3 (Fig. 1S – Appendix A). The optimal value *k*=2 is followed by *k*=4 (Δ*k*=7.62) and *k*=5 (Δ*k*=7.66) (Fig. 1S – Appendix A). For this reason, the individual Q-matrix barplot for k=2-5 is shown in Figure 2.3: for increasing values of *k*, the clustering of individuals in population 19 to form a separate group is still evident, and a more subtle structure involving the other populations is suggested (see particularly Fig. 2.3 *k*=5). This was, however, very weak as indicated by high level of admixture (i.e. vertical bars with more than one colours).

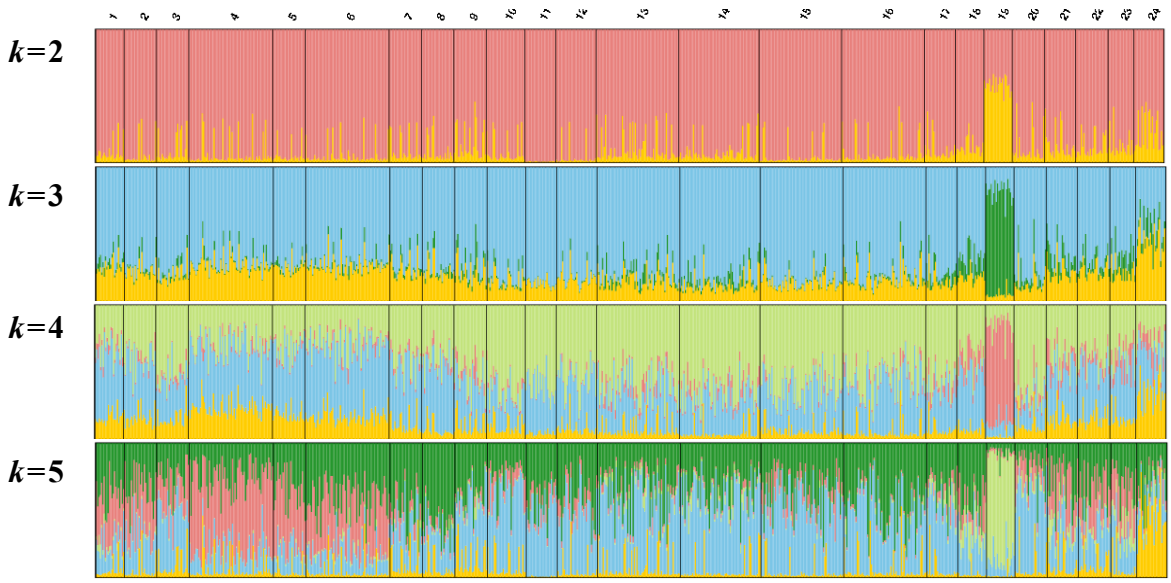

Fig 2.3. Clustering analysis by STRUCTURE v.2.3: bar-plots of individual Q-matrix for *k*=2-5.

To validate the results obtained with STRUCTURE v.2.3, a multivariate DAPC analysis (Jombart *et al.* 2010) was conducted on the same dataset. In successive *k-means* clustering, the minimum BIC values were reached for $k = 3$ (2732.568) and $k = 4$ (2732.491) (Fig. 2S – Appendix A). The results from the DAPC approach suggest the presence of four genetic pools of origin through the entire sample, while confirming a high

level of admixture within each sampled population, with the exception of population 19 (Fig. 2.4).

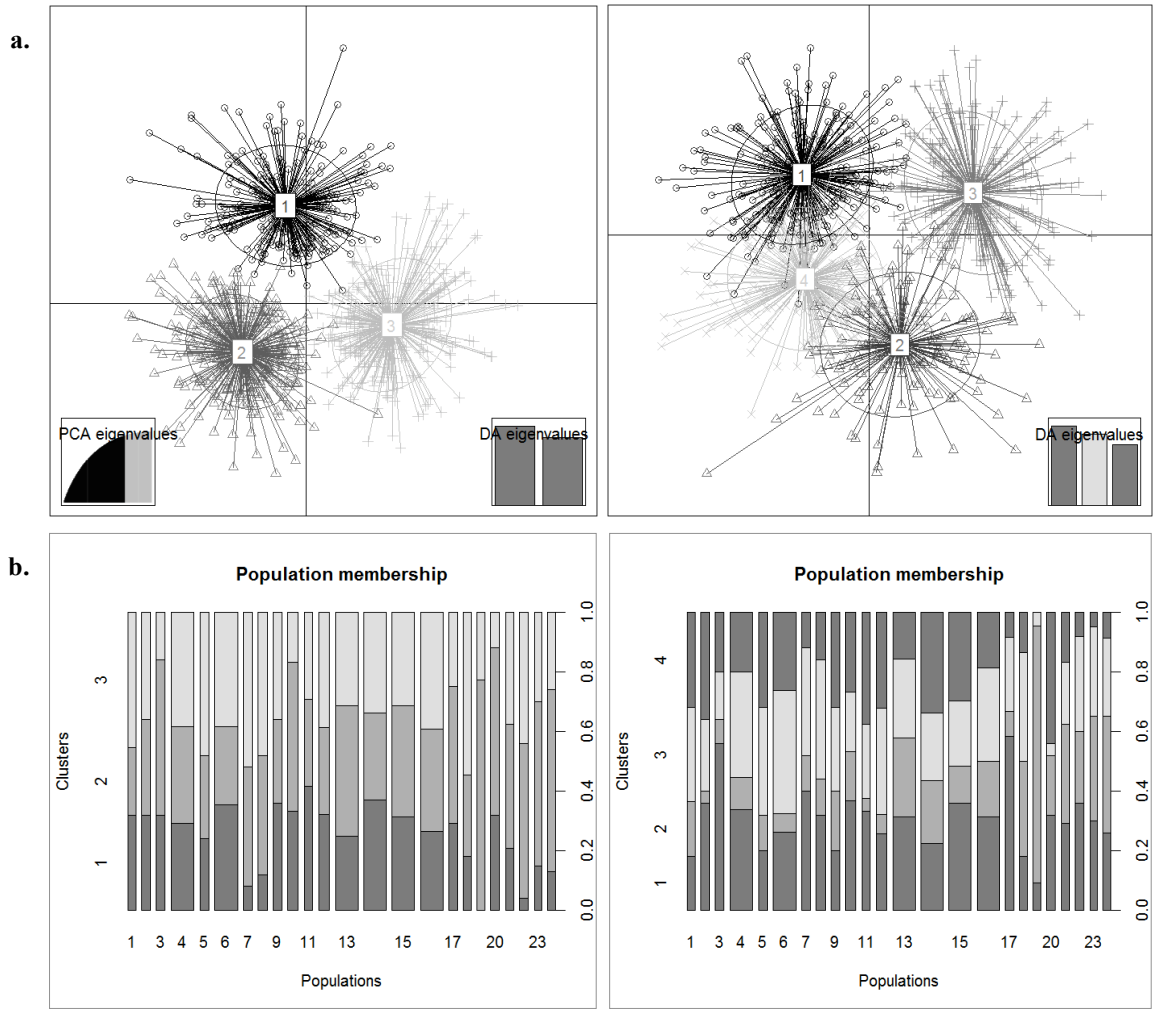

Fig. 2.4. (a.) Genetic clusters defined by DAPC analysis are represented by scatterplots: $k = 3$ (BIC = 2732.568 on the left) and $k = 4$ (BIC = 2732.491 on the right). **(b.)** For each k population membership to a genetic pool is showed by barplots. For both $k = 3$ (left) and $k = 4$ (right), population 19 belong mainly to a single cluster.

Population 19 clearly diverged from the others also when the sampled populations were used as prior groups in the DAPC analysis (Fig. 3S – Appendix A). Hence, although the two different approaches gave different answers as to the optimal number of original genetic clusters, they both revealed high levels of admixture within each sampled population and strong divergence of population 19, where most individuals belong to just

one cluster (Fig. 2.3, 2.4 and 3S – appendix A). As $k = 2$ better explains the distribution of genetic differentiation among populations (see also pair-wise F_{ST} estimates), it was chosen as the best number of clusters.

2.3.3 Climatic data

For each sampling site, eleven climatic indicators were derived: mean annual temperature (MAT), annual precipitation (AP), maximum temperature of warmest quarter (maxWmQT), minimum temperature of coldest quarter (minCQT), mean temperature of wettest and driest quarters (MWtQT and MDQT respectively) precipitation of wettest, driest, coldest and warmest quarters (WtQP, DQP, CQP and WmQP respectively) and growing degree days base 5°C (GDD5) (Table S5 – Appendix A). GDD5 is a temperature-based indicator, which refers to accumulation throughout the year of daily mean temperature values over a baseline that is the minimum temperature for growth. In boreal evergreen conifers from the northern hemisphere, such as species of *Picea*, this baseline temperature is 5°C (Prentice *et al.* 1992).

Using a two-sample Kolmogorov-Smirnov test, no significant difference between the annual precipitation distributions of the 10-year and 30-year trends was found (Table S6 – Appendix A), hence aligned time series for temperature and precipitation data could be used. However, as already stated by Hancock *et al.* (2008), it is important to note that the long-term climatic conditions experienced by each population may only be partially reflected by the values used. Estimates of correlations between climatic variables are shown in Table S7 (Appendix A).

2.3.4 F_{ST} outlier analysis

The program BayeScan v2.1 (Fall & Gaggiotti 2008) was used to analyze both the entire and reduced (no population 19) datasets to detect outlier SNPs. On the overall five F_{ST} outliers SNPs were identified at the 5% significance level (corrected FDR test; Fig. 2.5 a, b), while no outliers were found at the 1% significance level. When the analysis was performed on the entire dataset, F_{ST} outliers were identified for SNPs CL1852Contig1_01-81, 2_8852_01-97, 0_17587_01-42, 0_17587_01-392, all having low F_{ST} values, and therefore putatively under purifying or balancing selection (Table 2.4).

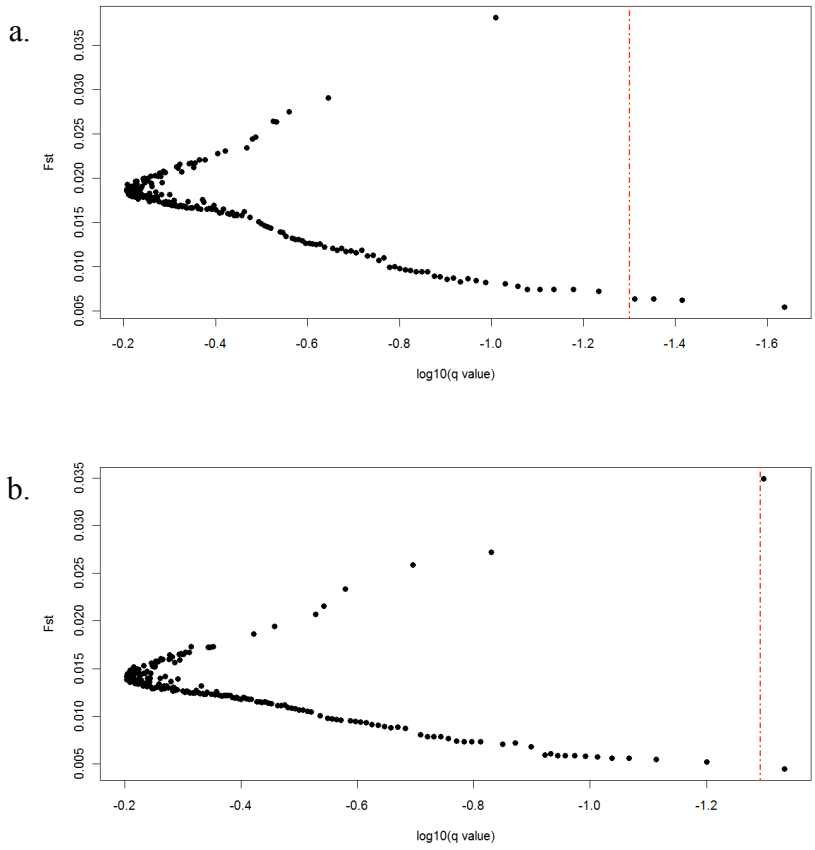

Fig. 2.5. F_{ST} outliers analysis using BayeScan (Fall & Gaggiotti, 2008) performed on (a.) full dataset and (b.) reduced dataset (23 pops.) Estimates of F_{ST} are plotted against the logarithm of q-value, which is the minimum False Discovery Rate at which a given locus may become significant.

After removing population 19 only two F_{ST} outliers were found for SNPs CL1852Contig1_01-81 and 2_5483_02-109 (Fig 2.5b, Table 2.4). The latter one was not found in the full dataset analysis, and was the only one with high F_{ST} value, hence putatively under positive selection.

Putative function for each locus was obtained by *blastx* search, looking for the best and more informative hits (see also supplementary material Table S2 – Appendix A). Of the five outlier loci detected, three resulted to be anonymous, while locus CL1852Contig1 encodes for a putative protein containing the poly-adenilate binding domain - usually found in proteins involved in post-transcriptional gene expression processes - and locus 2_8852 encodes for a putative galactokinase.

2.3.5 Environmental association analysis

The Bayesian generalized linear mixed model developed by Coop *et al.* (2010) was used to analyse the association between allele frequencies and climatic variables on the whole dataset, while taking into account the population structure. Hierarchical clustering of row and columns of the covariance matrix, determined in the first step of the model, identified population groupings largely corresponding to the results obtained in the previous population structure analysis (Fig. 4S – Appendix A). Eleven climatic variables were tested in the model. Bayes Factor (BF) values are showed in Table 2.4 and indicate "substantial" (average BF = 4.49) evidence for association to one, and up to four, variables in seven SNPs. Four SNPs were correlated with at least one precipitation variable (UMN_1604_01-348, CL1343Contig1_05-165, 2_374_01-319 and 0_18261_01-105). Two SNPs were associated to temperature-related variables (0_7471_01-399, 0_16480_02-185),

but still indirectly related to precipitation: mean temperature of the wettest and driest quarters.

Finally, one SNP (CL304Contig1_01-118) showed substantial association with both temperature and precipitation variables: mean temperature of the wettest and driest quarters and precipitation of the wettest and coldest quarters. The putative function was obtained for five of the associated loci, while two were anonymous loci. Locus UMN_1604 encodes for a sequence similar to SNF2 proteins family; locus CL304Contig1 encodes for a putative conserved domain: the oxygen-evolving enhancer protein1, which is involved in the photosystem II stabilization, where water oxidation occurs during photosynthesis; locus CL1343Contig1 encodes for a protein similar to phosphoenolpyruvate carboxykinase 2, critical enzyme involved in the gluconeogenesis; locus 0_18261 encodes for a sequence that has high similarity with the NAD(P)-linked oxidoreductase-like protein in *Arabidopsis thaliana*, which is involved in oxidation reduction on chloroplast thylakoid membrane, and finally locus 0_16480 encodes for a putative uncharacterized protein of unknown function.

Table 2.4. Summary of the SNPs detected using Bayenv and Bayescan methods. Associated climatic variables are reported according to their coding (see main text). Hypothetical function was defined according Best BLAST hit. Results are reported for both analyses on the full and reduced data-set. BayeScan values are in bold for significant outlier loci.

SNP		**Bayenv**		**BayeScan**			**Annotation**
Full data-set	Red. data-set	BF	Variables*	q-val	alpha	Fst	
UMN_1604_01-348[ns]	UMN_1604_01-348	3<BF<10	*AP*, WtQP, *DQP*, *CQP*	0.419 0.489[†]	0.135 0.124[†]	0.022 0.017[†]	SNF2 family DNA-dependent ATPase [*Physcomitrella patens subsp.patens*]
CL304Contig1_01-118[ns]		3<BF<10	MWtQT,MDQT,WtQP, CQP	0.477	0.128	0.022	PREDICTED: oxygen-evolving enhancer protein 1,chloroplastic-like isoform 2 [*Brachypodium distachyon*]
CL1852Contig1_01-81	CL1852Contig1_01-81	-		**0.023;** **0.046**[†]	**-1.459;** **-1.375**[†]	**0.005;** **0.005**[†]	Hypothetical protein containing the poly-adenilate binding domain. [*Populus trichocarpa*]
CL1343Contig1_05-165[ns]	CL1343Contig1_05-165	3<BF<10	*WmQP*	0.275 0.148[†]	0.347 0.608[†]	0.028 0.027[†]	phosphoenolpyruvate carboxykinase 2 [*Arabidopsis thaliana*]
2_8852_01-97[syn]		-		**0.038**	**-1.332**	**0.006**	PREDICTED: galactokinase [*Vitis vinifera*]
2_374_01-319	2_374_01-319	3<BF<10	*WmQP*	0.518 0.505[†]	0.094 0.131[†]	0.021 0.017[†]	-
0_7471_01-399		3<BF<10	MWtQT	0.226	0.410	0.029	-
0_18261_01-105		3<BF<10	WmQP	0.531	0.062	0.020	NAD(P)-linked oxiduriductase-like protein [*Arabidopsis thaliana*]
0_17587_01-42		-		**0.044**	**-1.315**	**0.006**	-
0_17587_01-392		-		**0.049**	**-1.313**	**0.006**	-
0_16480_02-185		3<BF<10	MWtQT, MDQT	0.619	0.032	0.019	uncharacterized protein LOC100277866 [*Zea mays*]
	2_5483_02-109	-		**0.050**[†]	**0.893**[†]	**0.035**[†]	-

[†] Values estimated when the analysis was performed on the reduced dataset. * Variables in *Italics* resulted as associated in both analyses on the full and reduced data-set. [ns] Non synonymous locus; [syn] Synonymous locus.

2.4 Discussion

This study presents a candidate gene-based approach to explore adaptive genetic variants in a non-model species of great relevance from both an ecological and economical point of view. Temperature and precipitation are among the major ecological variables that determine the natural distribution of plants and drive their adaptation (Berry & Björkman 1980; Manel *et al.* 2012), thus the main aim of our study was to identify potentially adaptive loci associated to these parameters in *P. abies*. Two distinct approaches were applied in order to first search for loci under selection, and then test for patterns of adaptive differentiation by detecting SNPs correlated to climatic variables.

On the overall, the combined approaches identified a set of 12 potentially adaptive loci, six of which fall in coding regions. When the Bayesian generalized linear mixed model (Bayenv) was used to test for correlation between SNP allele frequency and climatic variables, seven SNPs revealed to be associated to seasonal precipitations and seasonal temperatures. While all precipitation-derived variables had at least one SNP associated, only two temperature-derived variables were highlighted by the analysis: the mean temperature of the wettest quarter (MTWtQ) and the mean temperature of the driest quarter (MTDQ). Therefore, precipitation appears as the most involved variable in the SNP-climate relationship confirming the high sensibility of Norway spruce to soil water supply (Karlsson *et al.* 1997; Ditmarová *et al.* 2010).

The putative functions of these loci may provide information on their potential role in the adaptation process or response mechanisms to climatic changes. According to the best BLAST hit, five of the detected loci are putatively involved in regulation pathways or at least predicted proteins. For example, locus CL304Contig1 and locus 0_18261 are involved in mechanisms active during photosynthesis, therefore strictly affected by the

amount of light, temperature, and availability of water. In particular locus CL304Contig1 showed a similarity to a predicted oxygen-evolving enhancer protein 1, a catalytic site within photosystem II (PSII), which is responsible of water oxidation process (Ferreira *et al.* 2004). The SNP on this locus resulted associated to temperature and precipitation of wet and dry seasons. Several *in vivo* studies attested the damage to the PSII due to water stress (Canaani *et al.* 1986; Toivonen & Vidaver 1988), but more interesting, a study on wheat plants (Lu & Zhang 1999) demonstrated the existence of an antagonism between water stress and heat stress, with water stress enhancing thermostability of the oxygen-evolving complex under heat stress condition, thus possibly increasing the resistance of the whole plant to high temperature. The importance of temperature and water availability in the photosynthetic activity and efficiency is well established (Rennenberg *et al.* 2006) and evidence of adaptation at different temperature regime by photosynthetic rates was presented for conifer species (Rook 1969; Berry & Björkman 1980). This may well explain their association to temperature of wettest and driest quarters and to precipitation-derived variables. The association of locus CL1343Contig1, putatively involved in the gluconeogenesis, suggests that the amount of water during the warmest season could affect photosynthetic rate and hence the ATP production, leading to higher or lower rate of gluconeogenesis. Less clear deduction of the effects of climate can be driven from those loci involved in regulation of transcriptional processes (locus UMN_1604), although it is plausible to think of some abiotic stress-response processes. When the divergent population was removed from the analyses only three associated SNPs and two outlier SNPs were detected. This suggests that undetected loci may reflect selective constraints related to local climate for the removed population only.

Nevertheless, none of the SNPs that resulted associated to the climatic variables investigated was identified as outlier by BayeScan. At the 5% significant levels of posterior probability (corrected by FDR in ByeScan), the observed proportion of outlier loci (ranging from 1.8% to 2.8%) was comparable or a bit lower to rates reported with AFLPs in Norway spruce (Acheré *et al.* 2005), and positive selection was recorded only when F_{ST} outliers identification was conducted on the reduced dataset (removal of the divergent population). Between the five detected F_{ST} outlier, locus 2_8852 is a putative galactokinase and locus CL1852Contig1 is putatively involved in a post-transcriptional process.

According to Le Corre & Kremer (2012) in some cases F_{ST}-based methods may fail to identify the genetic causes of adaptation, especially for climate related traits in forest species, where loci detected as outliers seem to harbour low differentiation levels close to those measured at neutral markers. Such phenomenon would be even more evident if the trait is controlled by QTLs, hence the markers assessed may be in incomplete linkage with the true QTL. Although less likely, the mainly substantial (highest BF = 7.82) association estimated by Bayenv suggests a still weak signature of selection at those loci, maybe too recent for fixation to have eventually occurred (Le Corre & Kremer 2012; Hohenlohe *et al.* 2010), and thus less detectable. Therefore, the information brought by those loci associated to climatic variables should be considered as valuable as for the outlier loci, either because they are direct targets of selection or because they are genetically linked to a selected locus (Bonin *et al.* 2006).

Population 19, in the south-western Alps, was the most differentiated and divergent. Although previous findings, obtained using different markers, showed higher rate of genetic differentiation among Norway spruce population on the Italian Alps (F_{ST} = 0.118,

Scotti *et al.* 2000; F_{ST} = 0.05, Meloni *et al.* 2007), the overall genetic differentiation here detected by F_{ST}-*multilocus* (0.012) was comparable with the estimates obtained across the Alpine domain through analysis of sequence data (Heuertz *et al.* 2006) and within the standard range for conifer species, which maintain most of their variation within populations (Hamrick *et al.* 1992). The presence of a weak population structure in our sample was supported by the results obtained from both the Bayesian (STRUCTURE) and multivariate (DAPC) approach used. In fact, the multivariate method identified 4 genetic clusters, but highly admixed across all sampled populations with the only exception of population 19. Eventually, we can assert that populations from the eastern Italian Alps form one unique genetic cluster, which includes the isolated population on the Apennine; while the southernmost population of the western Alps represents a disjoint cluster. F_{ST} values in pair-wise comparison also reflect this pattern of genetic differentiation. Our results support the findings of Scotti *et al.* (2000). In their study on postglacial recolonization routes for Norway spruce in Italy, with SCAR markers, they asserted the uniqueness of the Valdieri population (our pop. 19), and excluded the existence of a glacial refugium represented by the Campolino population in the Apennines, which was instead grouped with eastern populations. In this study, given the high number of markers used, we provide a strong support for the hypothesis of Valdieri being a relict and isolated population.

To avoid false-positive associations detection it is necessary to account for demography and population structure (Excoffier *et al.* 2009), especially in species for which population structure covaries with climate (Holliday *et al.* 2010). Our sampling did not seem to match the extreme population structure that might strongly bias the statistics used to search for SNPs under selection. Therefore, no confounding effect due to

population structure should have affected our analyses. However, the differentiation of population 19 needed to be accounted for, and this was done by performing the analysis with and without this population, even when the software corrects for background levels of population structure (Bayenv).

Our sampling scheme tried to account for the natural distribution of the species across the Italian range, therefore the Alps, and the only Apennine population of Campolino. Although some limitations need to be acknowledged for what concern the sampling size in the western Alps, all populations were selected to maximize the variance of environmental conditions at sampling locations. All SNPs that were successful for genotyping and quality control were included in the analysis and scanned for adaptive value, this way the quality of the data was insured as well as the presence of putative candidate genes for climatic adaptation was inferred.

In conclusion, this study investigated the molecular basis of adaptation in a non-model species, of great economical and ecological interest at the European level, and that is acquiring growing interest in view of its full genome sequencing. Standing adaptive population genetics is a basic mean for natural plant population to respond to and track climatic changes (Barret & Schluter 2008; Holliday *et al.* 2010). Therefore this study, beyond assessing the transferability of markers identified in a *Pine* species to a different genus, provided a number of putative candidate loci, the role of which will be worth to further investigate, in order to obtain a valid tool for the identification of the adaptive potential of our forest. Moreover, the identification of potential candidate loci, which are meaningful in the adaptation process, and the introduction of rare but desirable alleles from natural population pave the way to future association studies (Wu 2002; Burdon & Wilcox 2007, Holliday *et al.* 2010). Furthermore, although beyond the main aim of simply evaluate

the genetic structure of sampled populations, the present work allowed for a more robust proof of the existence of a true relict population in the Maritime Alps (Valdieri), hypothesis previously proposed by Scotti *et al.* (2000) and Meloni *et al.* (2007), but still in need of further investigation.

Acknowledgments

First of all I would like to thank the co-authors of this work for their precious advices and comments to the manuscript, in alphabetical order: Giorgio Binelli, Nicola La Porta, Elena Mosca, David B.Neale, Duccio Rocchini

I would like to thank Dr Marco Pietrogiovanna of the Forest Service of the Province of Bozen, Dr. Raffaella Pettina of the Forest Service of the Province of Pistoia, the Forest Service of the Provinces of Trento and Udine, Dr Alessandro Maccabelli, David Blanco, Yuri Gori, Stefano Maffei, Elena Mosca, Marta Scalfi and Daniele Sebastiani for their support during the sampling. I am grateful to David Neale's group for the support and help in this work and during my staying at UCDavis, in particular I thank Katie Tsang and Randi Famula for the laboratory work in the DNA extraction, Ben Figurea for managing data storage in the database, Vanessa K. Rashbrook for their support in the genotyping. A particular acknowledgment is for Jill L. Wegrzyn and John D. Liechty for their help in the genotyping design and processing. Many thanks also to Andrew J. Eckert for his advice on the data analysis. I thank Luca Delucchi, Markus Metz and Markus Neteler for providing the climatic data and for their help.

This work and my PhD grant were supported by the ACE-SAP project c/o the Edmund Mach Foundation – San Michele all'Adige (TN), and partially funded by the Autonomous Province of Trento (Italy), with the regulation No. 23, June 12, 2008, of the University and Scientific Research Service.

References

Acheré V, Favre JM, Besnard G, Jeandroz S (2005) Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). *Molecular Ecology*, **14**, 3191–3201.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in ecology & evolution*, **23**, 38–44.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. *Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).*

Beniston M, Diaz HF, Bradley RS (1997) Climatic change at high elevation sites: an overview. *Climatic Change*, **36**, 233–251.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Bergmann F, Ruetz W (1991) Isozyme genetic variation and heterozygosity in random tree samples and selected orchard clones from the same *Norway spruce* populations. *Forest Ecology and Management*, **46**, 39–47.

Berry J, Bjorkman O (1980) Photosynthetic Response and Adaptation to Temperature in Higher Plants. *Annual Review of Plant Physiology*, **31**, 491–543.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular biology and evolution*, **23**, 773–783.

Borghetti M, Giannini R, Menozzi P (1988) Geographic variation in cones of Norway spruce (*Picea abies* (L.) Karst.). *Silvae genetica*, **37**, 178–184.

Bucci G, Vendramin GG (2000) Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. *Molecular ecology*, **9**, 923–934.

Buckley J, Butlin RK, Bridle JR (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology*, **21**, 267–280.

Burdon R, Wilcox P (2007) Population management: potential impacts of advances in genomics. *New Forests*, **34**, 187–206.

Canaani O, Havaux M, Malkin S (1986) Hydroxylamine, hydrazine and methylamine donate electrons to the photooxidizing side of Photosystem II in leaves inhibited in oxygen evolution due to water stress. *Biochimica et Biophysica Acta - Bioenergetics*, **851**, 151–155.

Chen J, Källman T, Ma X *et al.* (2012) Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics*, **191**, 865–881.

Coop G, Witonsky D, Rienzo AD, Pritchard JK (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, **185**, 1411–1423.

Ditmarová L, Kurjak D, Palmroth S, Kmet J, Strelcová K (2010) Physiological responses of Norway spruce (*Picea abies*) seedlings to drought stress. *Tree physiology*, **30**, 205–213.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eckert AJ, Bower AD, González-Martínez SC *et al.* (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular ecology*, **19**, 3789–3805.

Eckert AJ, Bower AD, Wegrzyn JL *et al.* (2009) Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics*, **182**, 1289–1302.

Endler JA (1986) *Natural Selection in the Wild. (MPB-21)*. Princeton University Press.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, **14**, 2611–2620.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Ferreira KN, Iverson TM, Maghlaoui K, Barber J, Iwata S (2004) Architecture of the photosynthetic oxygen-evolving center. *Science*, **303**, 1831–1838.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Frei E, Bodin J, Walther G-R (2010) Plant species' range shifts in mountainous areas—all uphill from here? *Botanica Helvetica*, **120**, 117–128.

Giannini R, Morgante M, Vendramin GG (1991) Allozyme variation in Italian populations of *Picea abies* (L.) Karst. *Silvae genetica*, **40**, 160–166.

Gömöry D, Longauer R, Hlásny T *et al.* (2011) Adaptation to common optimum in different populations of Norway spruce (*Picea abies* Karst.). *European Journal of Forest Research*, **131**, 401–411.

González-Martínez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. *The New phytologist*, **170**, 227–238.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A (2010) Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 2459–2468.

Hancock AM, Witonsky DB, Gordon AS *et al.* (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS genetics*, **4**, e32.

Haylock MR, Hofstra N, Tank AMGK *et al.* (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, **113**, D20119.

Heuertz M, De Paoli E, Källman T *et al.* (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics*, **174**, 2095–2105.

Hickling R, Roy DB, Hill JK, Fox R, Thomas CD (2006) The distributions of a wide range of taxonomic groups are expanding polewards. *Global Change Biology*, **12**, 450–455.

Hohenlohe PA, Phillips PC, Cresko WA (2010) Using polulation genomics to detect selection in natural populations:key concepts and methodological considerations. *International journal of plant sciences*, **171**, 1059–1071.

Holliday JA, Ritland K, Aitken SN (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *The New phytologist*, **188**, 501–514.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular ecology resources*, **9**, 1322–1332.

Huntley B, Birks HJB (1983) *An Atlas of Past and Present Pollen Maps for Europe, 0-13,000 Years Ago*. Cambridge [Cambridgeshire]: Cambridge University Press.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

IPCC (Intergovernmental Panel on Climate Change) (2007) *Summary for policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller (eds.)]*. Cambridge University Press, Cambridge, UK, and New York.

Jaillon O, Aury J-M, Noel B *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford, England)*, **23**, 1801–1806.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, **11**, 94.

Jump AS, Peñuelas J (2005) Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters*, **8**, 1010–1020.

Karlsson PE, Medin EL, Wallin G, Selldén G, Skärby L (1997) Effects of ozone and drought stress on the physiology and growth of two clones of Norway spruce (*Picea abies*). *The New Phytologist*, **136**, 265–275.

Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**, 773.

Lagercrantz U, Ryman N (1990) Genetic Structure of Norway Spruce (*Picea abies*): Concordance of Morphological and Allozymic Variation. *Evolution*, **44**, 38.

Latałowa M, van der Knaap WO (2006) Late Quaternary expansion of Norway spruce *Picea abies* (L.) Karst. in Europe according to pollen data. *Quaternary Science Reviews*, **25**, 2780–2805.

Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular ecology*, **21**, 1548–1566.

Li Y, Stocks M, Hemmilä S *et al.* (2010) Demographic histories of four spruce (*Picea*) species of the Qinghai-Tibetan Plateau and neighboring areas inferred from multiple nuclear loci. *Molecular biology and evolution*, **27**, 1001–1014.

Lin CH, Yeakley JM, McDaniel TK, Shen R (2009) Medium- to high-throughput SNP genotyping using VeraCode microbeads. *Methods in molecular biology*, **496**, 129–142.

Lu C, Zhang J (1999) Effects of water stress on photosystem II photochemistry and its thermostability in wheat plants. *Journal of Experimental Botany*, **50**, 1199–1206.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature reviews. Genetics*, **4**, 981–994.

Manel S, Gugerli F, Thuiller W *et al.* (2012) Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular ecology*, **21**, 3729–3738.

Mariac C, Jehin L, Saïdou A-A *et al.* (2011) Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. *Molecular ecology*, **20**, 80–91.

Meloni M, Perini D, Binelli G (2007) The distribution of genetic variation in Norway spruce (*Picea abies* Karst.) populations in the western Alps. *Journal of Biogeography*, **34**, 929–938.

Mosca E, Eckert AJ, Di Pierro EA *et al.* (2012) The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular ecology*, **21**, 5530–5545.

Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, Crassostrea virginica. *Molecular ecology*, **15**, 4229–4242.

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nature reviews. Genetics*, **12**, 111–122.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Neteler M (2005) Time series processing of MODIS satellite data for landscape epidemiological applications. *International Journal of Geoinformatics*, **1**, 133–138.

Neteler M (2010) Estimating Daily Land Surface Temperatures in Mountainous Environments by Reconstructed MODIS LST Data. *Remote Sensing*, **2**, 333–351.

Neteler M, Bowman MH, Landa M, Metz M (2012) GRASS GIS: A multi-purpose open source GIS. *Environmental Modelling & Software*, **31**, 124–130.

Nielsen R (2005) Molecular signatures of natural selection. *Annual review of genetics*, **39**, 197–218.

Paris M, Boyer S, Bonin A *et al.* (2010) Genome scan in the mosquito Aedes rusticus: population structure and detection of positive selection after insecticide treatment. *Molecular ecology*, **19**, 325–337.

Partanen J, Koski V, Hänninen H (1998) Effects of photoperiod and temperature on the timing of bud burst in Norway spruce (*Picea abies*). *Tree physiology*, **18**, 811–816.

Pavy N, Pelgas B, Beauseigle S *et al.* (2008) Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC genomics*, **9**, 21.

Petit RJ, Hampe A (2006) Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 187–214.

Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular ecology*, **19**, 2896–2907.

Prentice IC, Cramer W, Harrison SP *et al.* (1992) A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate. *Journal of Biogeography*, **19**, 117.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Quantum GIS Development Team (2011) *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project.

R Development Core Team (2009) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rennenberg H, Loreto F, Polle A *et al.* (2006) Physiological responses of forest trees to heat and drought. *Plant biology*, **8**, 556–571.

Rook DA (1969) The influence of growing temperature on photosynthesis and respiration of *Pinus radiata* seedlings. *New Zealand Journal of Botany*, **7**, 43–55.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*, **94**, 441–448.

Schmidt-Vogt H (1974) Das natürliche Verbreitungsgebiet der Fichte (*Picea abies* [L.] Karst) in Eurasien. *Allgemeine Forst- und Jagdzeitung*, 145:185–197.

Schmidt-Vogt H (1977) *Die Fichte*. Verlag Paul Parey, Hamburg, Germany.

Scotti I, Vendramin GG, Matteotti LS *et al.* (2000) Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified region (SCAR) markers. *Molecular ecology*, **9**, 699–708.

Shen R, Fan J-B, Campbell D *et al.* (2005) High-throughput SNP genotyping on universal bead arrays. *Mutation research*, **573**, 70–82.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular ecology*, **14**, 671–688.

Storz JF, Wheat CW (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution; international journal of organic evolution*, **64**, 2489–2509.

Sutinen R, Teirilä A, Pänttäjä M, Sutinen M-L (2002) Distribution and diversity of tree species with respect to soil electrical characteristics in Finnish Lapland. *Canadian Journal of Forest Research*, **32**, 1158–1170.

Theurillat JP, Guisan A (2001) Potential impact of climate change on vegetation in the European Alps: a review. *Climatic change*, **50**, 77–109.

Toivonen P, Vidaver W (1988) Variable Chlorophyll a Fluorescence and CO2 Uptake in Water-Stressed White Spruce Seedlings 1. *Plant Physiology*, **86**, 744–748.

Tollefsrud MM, Kissling R, Gugerli F *et al.* (2008) Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. *Molecular ecology*, **17**, 4134–4150.

Tuskan GA, Difazio S, Jansson S *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science (New York, N.Y.)*, **313**, 1596–1604.

Walther G-R, Post E, Convey P *et al.* (2002) Ecological responses to recent climate change. *Nature*, **416**, 389–395.

Wegrzyn JL, Lee JM, Liechty J, Neale DB (2009) PineSAP--sequence alignment and SNP identification pipeline. *Bioinformatics (Oxford, England)*, **25**, 2609–2610.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358.

Wu HX (2002) Study of Early Selection in tree breeding. 4. Efficiency of Marker-Aided Early Selection (MAES). *Silvae genetica*, **51**, 261–269.

CHAPTER 3

Adaptive variation in natural alpine populations of Norway spruce (*Picea abies* [L.] Karst) at regional scale: landscape features and altitudinal gradient effects.

Abstract

Steep environmental gradients represent interesting models to study the interaction between natural selection and gene flow, especially when aiming to better understand adaptation processes under global change. In the present study, the genetic basis of local adaptation along an altitudinal gradient were investigated in 18 natural alpine populations of Norway spruce (*Picea abies* [L.] Karst) sampled on a regional-scale on the Eastern Italian Alps, using Single Nucleotide Polymorphisms (SNPs). To account for patterns of gene flow and spatial genetic structure due to alpine landscape features, sampled populations were subdivided into three geographical groups, each including at least one sampled altitudinal gradient. Hierarchical analyses of molecular variance revealed that most of the genetic variability was found within populations (ca. 99%), and small but significant variation was also found among geographical groups (ca. 0.38%). In order to detect potentially adaptive markers, two distinct approaches were used. First, classical outlier detection was applied along the altitudinal gradients using both a Bayesian analysis

and a hierarchical island model, which showed contrasting results. Subsequently, Moran's eigenvector maps (MEM) variables, which may account for spatial variation and un-accounted environmental factors, were applied to an allele distribution model, and 19 loci significantly associated to environmental variation were identified. Four of these loci were also detected as outliers by the hierarchical island model. The combined approach of selection scan with a spatial analysis method allowed for a parallel investigation of potential candidate genes with environmental data, providing evidence for selective forces acting on loci of potential adaptive relevance in real landscapes.

Key words: Alpine landscape, local adaptation, Norway spruce, outlier loci, MEM, SNP.

3.1 Introduction

Pronounced environmental gradients and heterogeneous topography characterize the Alpine landscapes, where even small altitudinal changes can lead to large differences in temperature, humidity, exposure, and other environmental parameters (Körner 2003). Therefore, spatial variation in natural selection pattern can occur in this zone over a relatively small scale, leading to local adaptation and genetic differentiation between populations (Savolainen *et al.* 2007). According to the definition of Kawecki & Ebert (2004), local adaptation derives from the genotype-environment interaction describing allele's antagonistic effects on fitness in diversifying environments, where subsequent on-going natural selection creates the opportunity for adaptation in local populations. Spatial variation in the environment may be discrete, with several distinct habitat types, or it may consist of continuous environmental gradients, where many biotic and abiotic constraints vary (Endler 1977; Kawecki & Ebert 2004). Steep environmental gradients represent

interesting models to study the interaction between gene flow and selection, which determines the extant of local adaptation (Savolainen *et al.* 2007). In particular, Alpine landscape features, such as steep valleys and high mountain ridges, can affect the degree of genetic connectivity among populations, thus shaping patterns of gene flow and affecting micro-evolutionary processes (Manel *et al.* 2003; Latch *et al.* 2011). Therefore, spatial genetic structure of populations should be taken into account when searching for adaptive traits.

Local adaptation studies on natural populations allow for testing evolutionary hypotheses about traits favoured by particular environmental factors (Kaweki & Ebert 2004). In the actual ecologic context of climatic changes associated with global warming (IPCC 2007), the highly heterogeneous Alpine habitats are more susceptible at fast rates of environmental change (Theurillat & Guisan, 2001; Jump & Peñuelas 2005; Savolainen *et al.* 2007); in such an environment, the potential of many species to migrate fast enough to track the higher rate environmental variation is weak, especially for sessile organisms such as plants, which may need to rapidly adapt in place (Saxe *et al.* 2001; Jump & Peñuelas 2005). Furthermore, alpine environments are dominated by forest ecosystems, thus representing a suitable region for investigating the adaptive potential of conifer plant species under global change. However, studies on the genetic basis of adaptation in natural environments are still at the beginning, especially for non-model species, and very little is known about how local adaptation of populations will interact with future changes in climate (Savolainen *et al.* 2007; Jump & Peñuelas 2005).

Different approaches can be applied to identify potentially candidate loci associated to environmental variation, the most extensively used is based on Wright's fixation index (F_{ST}) outlier tests, generally classified in i) tests based in island model, assuming a unique

migration pool (Beaumont & Nichols 1996; Beaumont & Balding 2004; Foll & Gaggiotti 2008) and ii) tests based on hierarchical model, where populations are assumed to be hierarchically structured and migration rates are different among demes within groups and between groups (Excoffier *et al.* 2009). Landscape genetics approaches are currently proposed as alternative (Manel *et al* 2010, 2012) or a valid complementary method to the former one. They are based on allele distribution models and try to identify molecular markers whose changes in allele frequencies are correlated with environmental factors potentially acting as selective pressures and enforcing directional natural selection (Holderegger *et al.* 2008; Holderegger *et al.* 2010).

Altitudinal gradients are of particular interest in studying climate change effects on forest, and allow including tree-line eco-tones, which are particularly sensitive to environmental variation, especially under global warming (Saxe 2001; Erschbamer *et al.* 2008). Besides being an easily accessible ecological parameter, elevation may represent a compound factor, since in temperate regions, correlations of altitude to climatic factors such as temperature, precipitation and radiation, were assessed (Körner 2003). This could allow for the altitudinal gradients to mimic temporal variation in climatic conditions on a local scale, where the confounding effect of demography could be weaker or absent (Holderegger *et al.* 2008). Many forest trees show latitudinal and altitudinal differentiation in adaptive traits, Rehfeldt (1989) showed highly complex patterns of variation in North American mountainous forests (*Pseudotsuga menziesii* var. *glauca*), that could be accounted for by latitude, longitude, altitude, and slope; also an altitudinal component was shown in *Quercus petraea* for the cline in bud phenology (Ducousso *et al.* 1996).

The present study aims at assessing the adaptive potential of natural populations of Norway spruce (*Picea abies* [L.] Karst), a non-model coniferous species widely distributed

in European alpine environments. Its natural distribution extends from the European Alps eastwards to the Eastern Siberia and from the Balkan Peninsula to Scandinavia (Schmidt-Vogt 1974) representing three main domains: the Alpine, Hercyno-Carpathian and Baltic-Nordic domains. *P. abies* is predominantly outcrossing, with both seeds and pollen dispersed by wind (Burczyk *et al.* 2004). The Italian distribution of natural stands of *P. abies* covers the entire Alpine range along a wide altitudinal belt, with natural forests generally found between ca. 1500 m a.s.l. up to tree line (Mencuccini *et al.* 1995), and includes a single stand located in the northern Apennines, near Campolino (Borghetti *et al.* 1988). A previous investigation on genetic diversity of *P. abies* populations, across its Italian range, with Single Nucleotide Polymorphisms (Di Pierro *et al. in preparation*) showed a genetic differentiation between the eastern and western Alps, supporting previous findings with different genetic markers (sequence-characterized amplified region: Scotti *et al.* 2000; mitochondrial tandem repeat: Gugerli *et al.* 2001). Eastern populations were showed to form a single genetic cluster, characterized by higher within-population variation. Moreover, a clear association of both temperature and precipitation with genetic variation was found (Scalfi *et al. in preparation;* Di Pierro *et al. in preparation*).

In the current study, the association between environmental factors and genetic variation was investigated across 18 Norway spruce natural populations, sampled on a regional scale on the Eastern Italian Alps. To evaluate a possible genetic differentiation due to landscape features, populations were sampled across three main sub-regions delimited by two major watersheds as borders. The role of altitude as potential adaptive determinants was then tested on replicated altitudinal gradients sampled across the entire region. In particular, 175 non-synonymous single nucleotide polymorphisms (SNPs) were analysed to: i) test isolation due to landscape features, ii) assess an altitudinal effect along

the gradient, iii) search for putative candidate loci influenced by selection, and iv) test for association between allelic distribution and environmental variables by including all unaccounted spatial variation (Manel *et al* 2010; 2012). The approach used in this study, based on the parallel use of outlier detection methods with environmental variation analysis, aims to provide a link between molecular markers related to selection and specific ecological factors.

3.2 Material and Methods

3.2.1 Sample collection

The sampling design was optimized to discover patterns of local adaptation and it was conducted at a *regional*-scale, in the eastern part of the Italian Alps included in the Trentino Alto-Adige region (Fig. 3.1). In order to capture the broadest spectrum of adaptive genetic diversity, 18 Norway spruce natural stands were selected according to specific environmental factors across the region. In particular, four altitudinal gradients (each represented by a low, medium and high elevation plot) were sampled, one on the east (E) and one on the west (W) side of the Adige river, and two at north (N), surrounded by the Adige and Isarco rivers, (Fig. 3.1). Moreover, for each sampled population, elevation (m), aspect (degree) and slope (degree) were taken into account (Table S1 – Appendix B). From 25 to 65 mature trees per population were sampled, chosen at a distance of at least 10 meters from each other, and geographical coordinates were recorded by GPS device. Fresh needles from a total of 687 individuals were used in this study.

3.2.2 DNA extraction and SNPs genotyping

Desiccated needle tissue (50 mg) was homogenized using five 32" stainless steel grinding balls and liquid nitrogen in an automated tissue homogenizer (Spe SamplePrep, Metuchen, NJ, USA). Genomic DNA extraction was carried out using the standard DNeasy Plant 96 Kit (Qiagen) protocol on a Liquid Handling machine (Eppendorf). DNA was then quantified using Quant-iT PicoGreen Assay Kit (Invitrogen) on a fluorescent plate reader (PerkinElmer). SNPs selection and genotyping are described in detail in Di Pierro *et al.* (*in preparation*). In brief, a subset of SNPs derived from the CRPS re-sequencing project (http://dendrome.ucdavis.edu/NealeLab/crsp/overview), and originating from putative candidate genes that might be involved in abiotic or biotic responses, was selected for genotyping. SNPs genotyping was carried out at the UC Davis Genome Center, Davis CA, USA (http://www.genomecenter.ucdavis.edu) using the Illumina BeadXpress platform with the GoldenGate Genotyping assay (Shen *et al.* 2005). Intensity data for each SNP were then quantified and matched to specific alleles using GenomeStudio V2009.1 (Illumina). $GenCall_{50}$, *GenTrain* and Call Rate were used to evaluate the accuracy and efficiency of SNP genotyping, and a set of summary statistics was calculated for each SNP to control for the quality of the data. In the present study, in order to consider potentially candidate and informative SNPs, synonymous loci were not taken into account. A total of 175 high-quality and polymorphic SNPs (Table S2 - Appendix B), selected in 144 EST unigenes, were used.

3.2.3 Genetic diversity and geographical isolation

Expected heterozygosity (H_E) (Nei 1978) and inbreeding coefficient (F_{IS}) were calculated for each population, and global test of Hardy-Weinberg disequilibrium was performed by the Markov Chain (MC) algorithm. All computations were performed using GENEPOP 4.0.10 (Rousset 2008) with default parameters used for the MC algorithm. The genetic structure of sampled populations was assessed in a previous study (Di Pierro *et al. in preparation*) where weak or null differentiation among populations was revealed.

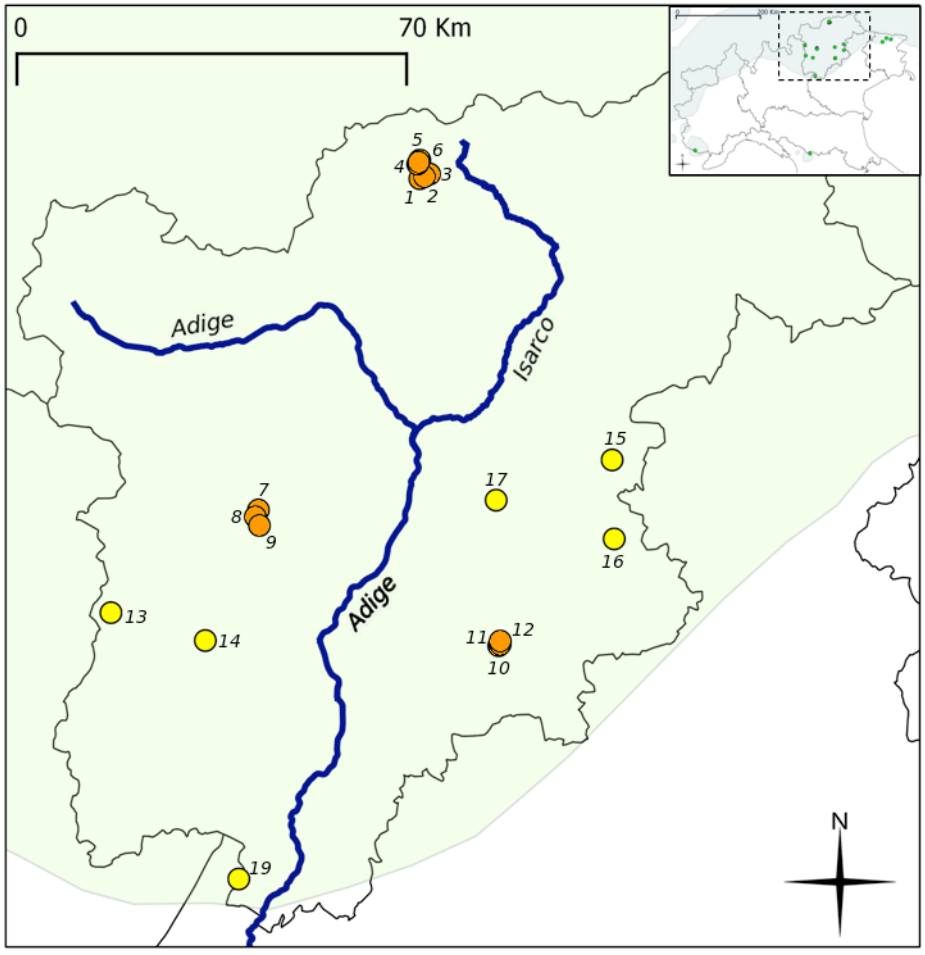

Fig. 3.1. Populations sampled in the Trentino Alto-Adige region. Adige and Isarco rivers were identified as major geographical borders to divide the region in three main sub-regions: East, West and North. At least one altitudinal gradient – highlighted in the map by three orange plots - is present in each sub-region. The map was realized using Qgis 7.1 software.

To understand whether and how landscape features and environmental gradients can shape genetic variation, three main geographical groups were identified using two major watersheds as borders: the Adige and Isarco rivers. Three sub-regions were so defined: North (N), East (E) and West (W) (Table S1 – Appendix B). To test for the topographic effects of mountainous landscape features, the proportion of genetic variance explained by these groups was tested in a hierarchical analysis of molecular variance (AMOVA; Weir & Cockerman 1984; Excoffier *et al* 1992; Weir 1996), using Arlequin version 3.5 (Ecoffier & Lischer 2010). Levels of significance were provided by 1,000 permutations.

3.2.4 Altitudinal gradient

Up to two altitudinal gradients were sampled in each sub-region, for a total of 12 populations distributed in four altitudinal gradients. A hierarchical AMOVA (Arlequin 3.5) was performed on the full SNPs set on the altitudinal plots only, to test for a possible effect of altitude. The hierarchic design grouped individuals within populations, within gradients and within geographical groups.

To assess the role of altitude as potential adaptive determinant, two different outlier loci methods were used to detect selection on SNPs along each altitudinal gradient. First, to account for the hierarchical levels described above, the hierarchical island model (Slatkin & Voelm 1991) implemented in Arlequin 3.5 (Excoffier & Lischer 2010) was used to perform coalescent simulations, deriving the null distributions of *F-statistics* and heterozygosities, from which locus-specific *p-values* were estimated. Then, a Bayesian approach for detecting outlier loci, implemented by the program BAYESCAN 2.1 (Foll & Gaggiotti 2008), was used across the SNPs dataset both on the whole group of 12 populations and by each of the 4 altitudinal gradients independently.

3.2.5 Moran's eigenvectors map variables and Environmental factors

To validate the previous analyses and to investigate in more details the effects of all the environmental gradients considered (elevation, aspect and slope), an approach that accounts for the spatial autocorrelation of samples was applied using all populations. As recently described by Manel *et al.* (2010, 2012), Moran's eigenvector map variables (MEMs) can be used to account for spatial and unmeasured environmental variables in regression analyses. For a given spatial weighting matrix, eigenvalues are equal to Moran's I coefficients of spatial autocorrelation and the associated eigenvectors represent the Moran's eigenvector map variables (Borcard & Legendre 2002; Dray *et al.* 2006). MEMs with positive autocorrelation (Moran's $I > 0$) describe global structures modelling broad-scale processes, while MEMs with negative autocorrelation (Moran's $I < 0$) describe local structure modelling spatial autocorrelation due to processes such as gene flow among sub-populations (Thioulouse *et al.* 1995; Dray *et al.* 2006; Borcard *et al.* 2011; Manel *et al.* 2012). In the present study, sampling sites' geographical coordinates were used to calculate the spatial weighting matrix, and MEM variables were computed using the method of principal coordinates of neighbour matrices (PCNM – Borcard & Legendre 2002; Dray *et al.* 2006; Borcard *et al.* 2011) with "PCNM" R package 2.12.2 (available at http://r-forge.r-project.org/R/?group_id=195). In order to explore the full spectrum of spatial variation, all MEM variables associated to significant Moran'I, together with the environmental factors obtained for each sampling sites (Table S1 – Appendix B), were used to detect loci of ecological relevance (Manel *et al.* 2012) and therefore to provide an indication of environmental parameters potentially involved in conifer plants adaptation.

Three environmental variables were obtained for each sampled populations using the GRASS GIS 6.4 (GRASS Development Team 2011, http://grass.osgeo.org; Neteler *et al.*

2012): elevation (m), slope (degree) and aspect (degree). To render the aspect variable appropriate for use in linear regression it was transformed in "folded aspect" about the north-south line prior to the analysis (McCune & Keon 2002) (Table S1 – Appendix B).

3.2.6 Regression analyses for the identification of putatively adaptive loci

First, a multiple regression analysis was performed with SNPs minor allele frequencies as dependent variables and MEMs, elevation, slope and aspect as independent variables. Subsequently, according to Manel *et al.* (2012), only significant loci after Holm correction for multiple testing, and having more than 50% of variation explained by the environmental and MEM variables ($R^2_{adj} \geq 0.5$), were referred to as *loci of ecological relevance* (hereafter denoted as *LER*). Their allele frequencies were tested for significant association with each environmental and MEM variable independently in a univariate linear regression model. This approach allowed accounting for potentially correlated variables.

3.3 Results

3.3.1 Genetic diversity and geographical isolation

Multilocus tests of heterozygosity deficiency were performed across all populations and results are presented in Table 3.1 together with F_{IS} and H_E estimates. Only two populations, number 15 and 18, revealed deviation from Hardy-Weinberg equilibrium, in both cases due to heterozygotes deficiency (*p-value* ≤ 0.05).

Table 3.1. Probability values of multi-locus heterozygosity deficiency test, performed across populations by the Markov Chain (MC) algorithm. Expected heterozygosity (H_E) and inbreeding coefficient (F_{IS}) values per population are reported.

Population	Locality†	H_E	F_{IS}	*p-value*
1	Val_Ridanna-l	0.272	-0.024	0.818
2	Val_Ridanna-m	0.264	0.022	0.273
3	Val_Ridanna-h	0.262	0.005	0.124
4	Val_Ridanna-l	0.267	-0.004	0.721
5	Val_Ridanna-m	0.261	-0.031	0.874
6	Val_Ridanna-h	0.265	-0.011	0.614
7	Val_Sole-l	0.258	-0.020	0.601
8	Val_Sole-m	0.251	0.016	0.234
9	Val_Sole-h	0.257	-0.012	0.591
10	Val_Calamento-l	0.265	-0.024	0.947
11	Val_Calamento-m	0.260	0.008	0.234
12	Val_Calamento-h	0.266	-0.017	0.364
13	Val_Genova	0.265	-0.013	0.845
14	Val_Algone	0.255	-0.001	0.303
15	Val_SanNicolo'	0.267	0.023	0.025*
16	Paneveggio	0.262	0.006	0.365
17	Passo_Lavaze'	0.262	-0.039	0.920
18	Avio	0.262	0.037	0.012*

†Elevation plots: l= low, m=medium, h= high; * *p-value* ≤ 0.05

The AMOVA performed to test genetic differentiation among the sub-regions E, W and N revealed a pattern of geographical isolation possibly due to alpine topography: the discontinuities due to the valleys of the Adige and Isarco rivers, in addition to distance, may have shaped the genetic diversity of populations among the three sub-regions. Indeed, highly significant genetic variation was present at all levels: among sub-regions, among populations within sub-regions, and within populations (Table 3.2). The variability within population accounts for the major percentage of variation (99.33%) confirming the high levels of intra-population genetic variation in conifer trees (Hamrick *et al*. 1992; Müller-Starck 1995; Abril *et al.* 2011).

Table 3.2. Analysis of Molecular Variance (AMOVA) was performed to assess isolation due to geography. Genetic differentiation was tested among sub-regions: East *vs* West *vs* North.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	*p-value*	Fixation Indices
Among Groups (F_{CT})	2	128.140	0.08076	0.38	0.000^{***}	0.004
Among populations within groups (F_{SC})	15	387.830	0.06183	0.29	0.000^{***}	0.003
Within Populations (F_{ST})	1356	28856.667	21.28073	99.33	0.000^{***}	0.007
Total	1373	29371.529	21.42332			

3.3.2 Altitudinal gradient and F_{ST} outliers detection

Highly significant differentiation among the altitudinal gradients was detected (*p-value* ≤ 0.001), attesting the effect of the geographical groups. The variation among populations within gradients was also highly significant, hence indicating a possible altitudinal effect on the genetic diversity of these plots, although the percentage of genetic diversity measured was very low (0.20%) (Table 3.3).

According to the results obtained by the two methods of outlier detection used, contrasting evidence of adaptation along the altitudinal gradients was found. The analysis on all 12 altitudinal plots with BAYESCAN identified one outlier locus (at the 5% significance level, corrected FDR test): 2_5483_02-109 (*qval* = 0.046; F_{ST} = 0.035). However, when BAYESCAN was used to analyse each gradient independently no F_{ST} outliers were detected. Conversely, the hierarchical approach performed in Arlequin reported 25 outlier loci with significant *p-value* (Fig. 3.2; Table S3 – Appendix B), including locus 2_5483_02-109 detected also by BAYESCAN. This locus was identified among the most significant by Arlequin's hierarchical method, together with locus

CL697Contig1_03-204. Out of the 25 identified loci, potential signature of positive selection was detected only for seven loci (2_5483_02-109, 0_13680_01-216, 2_5073_01-488, 0_9457_01-421, UMN_1604_01-348, 2_4281_02-253, CL1920Contig1_01-146), while most outliers showed lower differentiation than expected. False discovery rate (FDR) correction was subsequently applied to *p-values* of the outliers identified by Arlequin, to control for type error I. After correction for multiple tests, only locus CL697Contig1_03-204 was still significant (Table S3 – Appendix B).

Table 3.3. Analysis of Molecular Variance (AMOVA) was performed to test the effect of altitude. Genetic differentiation was tested within gradients within sub-regions.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	*p-value*	Fixation Indices
Among Groups (F_{CT})	3	123.514	0.09206	0.43	0.000***	0.004
Among populations within groups (F_{SC})	8	466.653	0.04238	0.20	0.000***	0.002
Within Populations (F_{ST})	762	28856.667	21.34345	99.37	0.000***	0.006
Total	773	29371.529	21.47789			

3.3.3 Identification of putatively adaptive loci and environmental predictors

The PCNM method computed ten Moran's eigenvectors map variables describing global (MEM1 to MEM4) and local (MEM5 to MEM10) structures. The six MEM variables associated to significant Moran's I (MEM1, MEM2, MEM7, MEM8, MEM9, and MEM10) (Fig. 1S – Appendix B) were used in the regression analyses.

The multiple linear regressions identified 19 loci significantly related to the environmental variables and MEM variables with $R^2_{adj} \geq 0.5$ (Table S4 – Appendix B),

which were subsequently regressed against each of the environmental variables and MEMs independently. The univariate regressions revealed both broad-scale and local-scale MEM variables significantly related to allele frequencies in 13 SNPs loci, but they were identified as major predictors (highest R^2_{adj} values) at six (UMN_1604_01-348, CL4023Contig1_01-114, CL1688Contig1_01-106, 2_6635_01-85, 2_4586_01-365, 0_13680_01-216) and four (CL1343Contig1_05-165, CL1224Contig1_01-546, 2_9845_01-282, 0_8844_01-281) loci, respectively (Table 3.4).

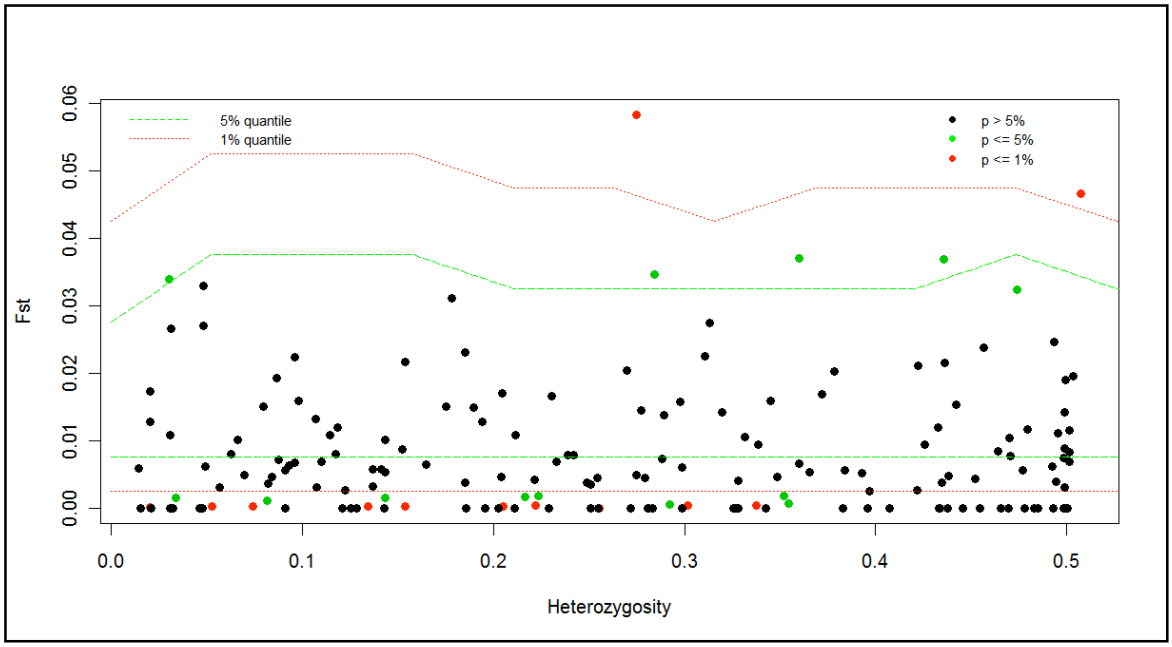

Fig. 3.2. Outlier loci detected by the Hierarchical Island Model implemented in Arlequin 3.5. The plot shows F_{ST} values, conditional on heterozygosity of the 175 SNPs loci investigated (See also values in Table S3 – Appendix B). Confidence interval limits obtained from simulated data are plotted as dashed lines. Loci significant at 5% (green dots) and 1% (red dots) are shown.

Broad-scale MEM variables appear to incorporate the spatial structure due to topography (Fig. 3.3), while less clear is the variation explained by local-scale MEM variables. Elevation, slope and aspect were the major factors affecting allele frequencies at five (CL1308Contig1_03-181, CL1225Contig1_03-91, 2_6313_01-164, 0_8531_01-363,

0_15075_01-341), two (CL1225Contig1_03-91, 2_1023_01-130) and three (CL4257Contig1_01-391, 2_4196_01-201, 0_7973_01-149) loci, respectively (Fig. 3.4). When a locus was significantly dependent from both environmental factors and MEM variables, the environmental factors were the major predictors (highest R^2_{adj} values; Table 3.4).

Table 3.4. Putatively adaptive loci and major predictors identified by the univariate regression. R^2_{adj} values are reported in the table for each significant[1] variable.

SNP loci	Elevation	Slope	Folded aspect	Broad-scale MEMs	Local-scale MEMs
UMN_1604_01-348				0.31**	0.24*
CL4257Contig1_01-391			0.25*		0.20*
CL4023Contig1_01-114				0.40**	
CL1688Contig1_01-106				0.50***	
CL1343Contig1_05-165				0.21*	0.26*
CL1308Contig1_03-181	0.32**			0.18*	
CL1225Contig1_03-91	0.23*	0.29*			
CL1224Contig1_01-546					0.46**
2_9845_01-282					0.20*
2_6635_01-85				0.48***	0.21*
2_6313_01-164	0.39**				
2_4586_01-365				0.25*	
2_4196_01-201			0.30*		
2_1023_01-130		0.20*			
0_8844_01-281				0.26*	0.31**
0_8531_01-363	0.16*				
0_7973_01-149			0.22*		
0_15075_01-341	0.42**				0.17*
0_13680_01-216				0.53***	

[1] *p-value*:≤ 0.001 ***; ≤ 0.01 **; ≤ 0.05 *

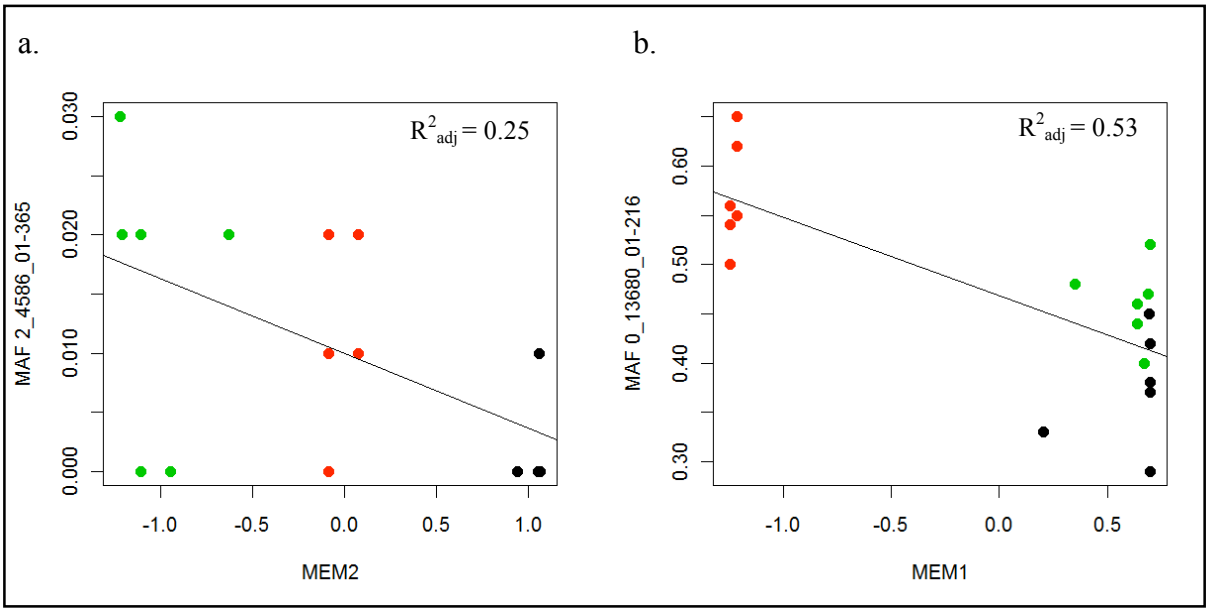

Fig. 3.3. Significant linear regression between SNP minor allele frequencies (MAF) and broad-scale Moran's eigenvectors map variables: a) MEM2 was significantly associated only to locus 2_4586_01-365; b) MAF of locus 0_13680_01-216 had the highest proportion of variation explained by MEM1. In both cases the association appears to describe the spatial structure due to topography. Different dot colours indicate populations of different geographic sub-regions. N = 'red', W = 'green', E = 'black'.

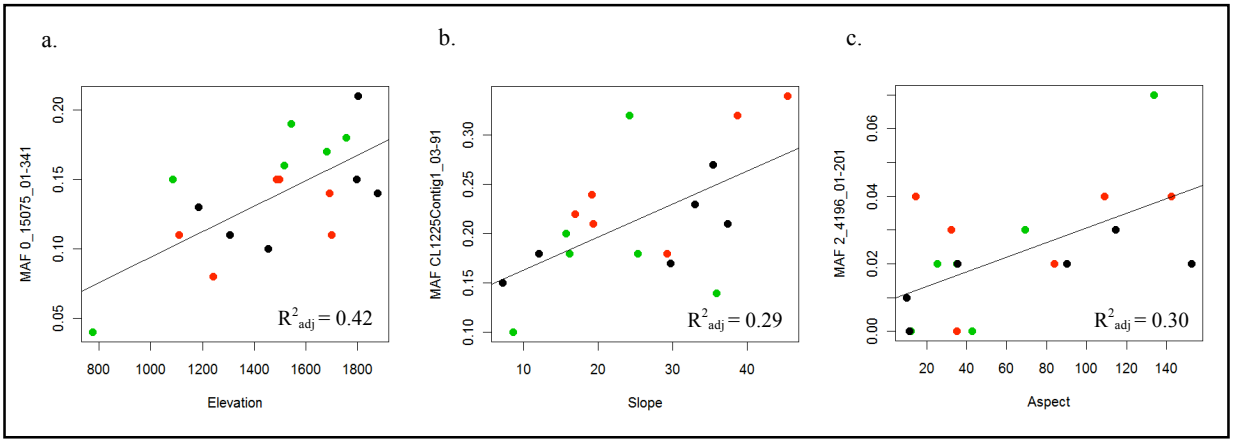

Fig. 3.4. Significant linear regressions between SNP minor allele frequencies and environmental factors: a) elevation, b) slope and c) aspect. Minor allele frequencies of those loci having the highest proportion of variation explained by the related environmental variables were plotted. Different dot colours indicate populations of different geographic sub-regions. N = 'red', W = 'green', E = 'black'.

3.4 Discussion

The present study investigated landscape elements and environmental factors in natural alpine populations of *Picea abies*, with the aim of studying adaptive genetic variation along an altitudinal gradient. All populations belonged to the same genetic cluster (Di Pierro *et al*. *in preparation*) and were sampled on a regional-scale.

3.4.1 Genetic differentiation and topographic effect

The identification of three main geographical sub-regions revealed a pattern of small but significant genetic differentiation between eastern, western and northern populations within the Trentino Alto Adige region, suggesting a spatial structuring likely due to the territory topography. The Alpine landscape, characterized by steep valleys and high mountain ridges, is likely to affect the degree of genetic connectivity among populations (Gugerli *et al.* 2001; Csaikl *et al.* 2002; Cain *et al.* 2000). Spatial genetic structure caused by restricted gene flow and leading to isolation patterns might considerably interfere with outlier detection and landscape genetics approaches (Holderegger *et al.* 2008; Excoffier *et al.* 2009; Holderegger *et al.* 2010). Therefore, the identified geographical isolation was taken into consideration when searching for local adaptation.

Altitudinal effect on genetic variation was tested on four replicated gradients by a hierarchical approach, to account for differentiation among the three sub-regions. AMOVAs' results indicated that elevation accounts for a low, but significant, partition of total genetic variation along each gradient. Numerous ecological and evolutionary factors, such as mating system, dispersal of pollen and seeds could determine differentiation along an altitudinal gradient. Potential altitude-dependent site variations could affect

phenological traits such as bud flushing or even pollen and seeds production, strictly correlated to temperature conditions (Sarvas 1968; Mencuccini *et al.* 1995). This phenomenon could lead to changes of the mating system (e.g. asynchrony between pollen release and ovule receptivity) and promotion of assortative mating. Moreover, the downward translocation of seeds in alpine regions, could promote a decrease in the "high-elevation alleles" frequency, if no strong selection acts (Geburek *et al.* 2007). Nonetheless, our results appear to contrast most previous findings, where lack of genetic differentiation due to elevation was observed, even along steep ecological gradients (Gugerli *et al.* 2001; Bingham & Ranker 2000; Unger *et al.* 2011). Indeed, only a few studies have demonstrated altitudinal differentiation (e.g. Bergmann 1978; Geburek *et al*. 2007) and usually when specific genes, which are either directly selective or markers tightly linked to such genes, are investigated. However, in our study the significant differentiation detected within gradients, by Arlequin's hierarchical AMOVAs, explained only a very low percentage of genetic variation (0.2 %), and it was described by a fixation index (F_{SC} = 0.002) comparable to what found by Unger *et al.* (2011). In their study, spatial genetic structure along an altitudinal transect was investigated and nearly identical results were found (F_{ST} = 0.002), indicating very small population differentiation. Therefore, our results are likely to suggest a possible selective effect on just a few SNPs rather than a real genetic differentiation due to altitudinal clines acting on many loci (Savolainen & Pyhäjärvi 2007; Hedrick 2006).

3.4.2 F_{ST} outlier detection along altitudinal gradients

Outlier loci potentially under selection along the gradient were thus searched and 25 putative adaptive SNPs were identified by hierarchical island model implemented in

Arlequin 3.5. Contrasting results were obtained using the Bayesian approach: no significant loci were found when the analysis was performed within each gradient, and only one locus (2_5483_02-109) was significantly detected when all the 12 altitudinal populations were pooled together. Discordant results among outlier methods have been recently assessed by Narum & Hess (2011) and have been observed in studies conducted on other species and using other type of markers (Tice & Carlon 2011; Vilas *et al.* 2012). Caution is needed when interpreting these results and particular control for Type error I is required. Among the F_{ST} outlier tests compared by Narum & Hess (2011) using SNPs, there were both Arlequin's hierarchical method and BAYESCAN, which was denoted as one of the most reliable methods to detect outliers from the distribution of locus-specific F_{ST}. However, the latter does not account for the hierarchical structure of populations while searching for outliers. In our study, although small, the identified differentiation between geographical sub-regions and along the altitudinal gradients was significant, thus both methods were used. Although 14% of the analysed loci were detected as outliers by the hierarchical island model, only seven loci showed potential signatures of positive selection, including locus 2_5483_02-109 (F_{ST} = 0.058), but the majority represented lower differentiation than expected under neutrality (see also Fig 3.2). To evaluate the role of Type error I and its effect on outlier detection, false discovery rate (FDR) correction was applied to the outliers identified by Arlequin, and only locus CL697Contig1_03-204 (F_{ST} = 0.00001) was finally detected as significant F_{ST} outlier. In the context of local adaptation studies, outlier loci are usually of great relevance because selection may be the underlying cause of their atypical behaviour, either because they are directly under selection or because they are genetically linked to a selected locus (Bonin *et al.* 2006). However, our results confirmed the necessity of caution when interpreting outliers in genome scans.

Different approaches not based on the categorisation of loci into outlier and non-outlier, which involves an arbitrary cut-off, may be advisable. Such an approach could be represented by models where natural selection along environmental gradients or heterogeneity is assumed to induce clinal variation in allele frequencies at loci linked to selected genes (i.e. allele distribution models) (Endler 1986; Joost *et al.* 2007; Holderegger *et al.* 2008; Manel *et al.* 2010), and applied by using regression analyses with specific environmental factors and gradients (Butlin 2010; Narum & Hess 2011).

3.4.3 Moran's eigenvectors map variables and Environmental factors

Appropriate characterization of the environmental variables potentially associated with allele frequencies across the genome is important for genetic association studies in natural populations, however a comprehensive collection of all environmental and habitat data involved may not always be possible (Narum *et al.* 2010). An appropriate method to incorporate the effect of un-accounted environmental variables is based on Moran's eigenvector maps variables (MEM) (Borcard & Legendre 2002) and has been recently proposed to investigate adaptive variation in different alpine plant species (Manel *et al.* 2012) and at different spatial scales (Manel *et al.* 2010). The same approach was applied in this study using broad-scale and local-scale MEM variables together with elevation, slope and aspect variables. Nineteen *LER* (Manel *et al.* 2010; Manel *et al.* 2012) were identified, and their major adaptive determinants were established. Elevation, slope and aspect variables influenced allele distribution at more than 20%, 10% and 15% of the detected *LER*, respectively, while more than half (53%) were significantly associated only to MEM variables, which represented un-accounted environmental variables and geographical variation.

3.4.4 Putative informative loci detection across methods

Four of the identified *LER* were also detected by Arlequin's hierarchical analysis: 0_13680_01-216, 2_9845_01-282, UMN_1604_01-348, significantly associated to broad, local and both types of MEM variables respectively, and SNP CL1225Contig1_03-91, significantly influenced by both elevation (R^2_{adj} = 0.23) and slope (R^2_{adj} = 0.29). No information about the SNP type is available for these loci, with the exception of UMN_1604_01-348, a non-synonymous SNP, whose locus (UMN_1604) encodes for a sequence similar to a helicase protein of SNF2 family. The proteins of this family are generally involved in several cellular processes such as transcriptional regulation, and in various stages of processing of DNA damage (Eisen *et al.* 1995). Putative functions for detected loci were obtained by BLAST search of the best informative hit in the GenBank - NCBI repository (Table S3 and Table S4 – AppendixB). The only locus detected by BAYESCAN (2_5483_02-109), was significantly associated (*p-value* = 0.006) to MEM1 by the multiple linear regression, however it was not included among the *LER* due to the low R^2_{adj} value (0.3176). Locus 2_5483_02-109, was detected by the same outlier method (BAYESCAN) in a previous investigation on *P.abies* with 24 sampled populations across the Italian Alps and using both synonymous and non-synonymous SNPs (Di Pierro et al. *in preparation*). No information is available on the putative function of this locus and BLAST search matched (e-value = 2.00e-94) only an unknown protein inferred from a cDNA sequence obtained from *Sitka spruce* buds (Gen Bank accession: BT123678). In the same previous study, UMN_1604_01-348 was highlighted as significantly associated to annual and seasonal precipitation.

Loci are present between detected *LER*, which were identified in previous studies as associated to climatic components or as outliers. Specifically, locus CL1688Contig1

(influenced by broad-scale MEMs) was found to be associated to climatic factors in *Pinus mugo* populations on the Italian Alps (Mosca *et al.* 2012) and in *Picea abies* populations sampled at European level (Scalfi *et al. in preparation*). This locus showed a similarity to a putative β-Galactosidase of *Vitis vinifera.* These types of proteins are involved in pectin degradation, which cause structural changes and modify cell wall porosity (Esteban *et al.* 2003). Mosca *et al.* (2012) have previously indicated as associated to principal components of climatic factors four more loci detected in this study. In particular, SNP 2_6313_01-164 here showed significant association to elevation, its putative gene product is similar to a conserved protein acting as regulator of mitotic phase during cellular cycle, and hypothetical in *Ricinus communis*; SNP 2_1023_01-130 was related to slope, it is a non-synonymous SNP, and its locus (2_1023) encodes for a protein similar to a putative ribosomal protein S2 in *Pinus lambertiana*; CL1224Contig1_01-546, here associated to a local-scale MEM, is present in a locus (CL1224Contig1) coding for a sequence similar to alpha-N-acetylglucosaminidase of *Medicago truncatula.* It has been recently showed (Ronceret *et al.* 2008) the important role played by this protein in plant reproductive stages during male gametogenesis and early seed development; finally, the non-synonymous outlier locus 0_10267_01-148 encodes for a putative protein similar to a transcriptional activator: myb domain protein.

Considering all detected outliers, including also the 25 outliers found by Arlequin's and significant for *p-value*, concordance between the two approaches used in this study was observed at only few loci. For example, the significant clinal pattern due to elevation observed at four SNPs (CL1308Contig1_03-181, 2_6313_01-164, 0_8531_01-363, 0_15075_01-341) was not highlighted by the F_{ST} outlier methods. In particular, 0_15075_01-341 is a non-synonymous SNP within a gene encoding for a putative protein

similar to GRX1-glutaredoxin. GRX proteins seem to be involved in regulation pathways of flowering time, root and shoot development (Xing *et al.* 2006), thus being possible candidates for displaying adaptive variation. On the other hand, additional loci were identified as candidates with the F_{ST} outlier approach but did not correlate with any of the variables tested in this study. An example is CL1920Contig1_01-46, a non-synonimous SNP showing high adaptive differentiation, and present in a region encoding for a protein similar to chitinase-like in *Vitis vinifera*. Besides playing a role in the defence of the organism against pathogen attack, chitinase activity or chitinase-genes transcription may be induced by external stimuli such as drought, cold, and UV light (Kasprzewska 2003).

3.4.5 Conclusion

The studied populations belonged all to a single genetic cluster, however landscape features were identified that may affect gene flow between them. This study could complement previous investigations at broader scales on Alpine conifers. Several studies have been conducted to explore environmental gradients on a broad scale (e.g. exploring latitudinal and longitudinal clines: Chen *et al.* 2012; Mosca *et al.* 2012), but only a few focused on local gradients such as altitude (Unger *et al.* 2011; Scalfi *et al. in preparation*).

Characterization of local adaptation with putative candidate markers may be a useful starting tool to determine adaptive variation in natural populations, especially for non-model species. Therefore the identified loci may represent a plausible target for further analysing the potential for evolutionary change in the current ecological context (Savolainen *et al.* 2007). Given the complexity of natural environments and that significant signal may also due to linkage with genes or markers not included in this study

(Charlesworth *et al.* 1997), validation of genetic and phenotypic associations under controlled environments should be an advisable future scope.

Acknowledgments

First of all I would like to thank the co-authors of this work for their advices and comments to the manuscript, in alphabetical order: Giorgio Binelli, Santíago C. González-Martínez, Nicola La Porta, Elena Mosca, David B.Neale.

I am grateful to Jill L. Wegrzyn and John D. Liechty for their help in the genotyping data re-processing. I thank Luca Delucchi, Markus Metz Markus Neteler and Duccio Rocchini for providing the environmental data and for their help.

This work and my PhD grant were supported by the ACE-SAP project c/o the Edmund Mach Foundation – San Michele all'Adige (TN), and partially funded by the Autonomous Province of Trento (Italy), with the regulation No. 23, June 12, 2008, of the University and Scientific Research Service.

References

Abril N, Gion J-M, Kerner R *et al.* (2011) Proteomics research on forest trees, the most recalcitrant and orphan plant species. *Phytochemistry*, **72**, 1219–1242.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **263**, 1619–1626.

Bergmann F (1978) The allelic distribution at an acid phosphatase locus in Norway spruce (*Picea abies*) along similar climatic gradients. *Theoretical and Applied Genetics*, **52**, 57–64.

Bingham R, Ranker T (2000) Genetic Diversity in Alpine and Foothill Populations of Campanula rotundifolia (*Campanulaceae*). *International journal of plant sciences*, **161**, 403–411.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular biology and evolution*, **23**, 773–783.

Borcard D, Gillet F, Legendre P (2011) *Numerical Ecology with R*.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Borghetti M (1988) Geographic variation in cones of Norway spruce (*Picea abies* (L.) Karst.). *Silvae genetica*, **37**, 178–184.

Burczyk J, Lewandowski A, Chalupka W (2004) Local pollen dispersal and distant gene flow in Norway spruce (*Picea abies* [L.] Karst.). *Forest Ecology and Management*, **197**, 39–48.

Butlin RK (2010) Population genomics and speciation. *Genetica*, **138**, 409–418.

Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American journal of botany*, **87**, 1217–1227.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical research*, **70**, 155–174.

Chen J, Källman T, Ma X *et al.* (2012) Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics*, **191**, 865–881.

Csaikl U, Burg K, Fineschi S *et al.* (2002) Chloroplast DNA variation of white oaks in the alpine region. *Forest Ecology and Management*, **156**, 131–145.

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.

Ducousso A, Guyon J, Krémer A (1996) Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt) Liebl). *Annals of Forest Science*, **53**, 775–782.

Eisen JA, Sweder KS, Hanawalt PC (1995) Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Research*, **23**, 2715–2723.

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press.

Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press.

Erschbamer B, Kiebacher T, Mallaun M, Unterluggauer P (2008) Short-term signals of climate change along an altitudinal gradient in the South Alps. *Plant Ecology*, **202**, 79–89.

Esteban R, Dopico B, Muñoz FJ *et al.* (2003) Cloning of a Cicer arietinum β-Galactosidase with Pectin-Degrading Function. *Plant and Cell Physiology*, **44**, 718–725.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, **10**, 564–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Geburek T, Robitschek K, Milasowszky N, Schadauer K (2007) Different cone colours pay off: lessons learnt from European larch (*Larix decidua*) and Norway spruce (*Picea abies*). *Canadian Journal of Botany*, **85**, 132–140.

Gugerli F, Sperisen C, Büchler U *et al.* (2001) Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. *Molecular ecology*, **10**, 1255–1263.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hedrick PW (2006) Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 67–93.

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in Plant Science*, **15**, 675–683.

Holderegger R, Herrmann D, Poncet B *et al.* (2008) Land ahead: using genome scans to identify molecular markers of adaptive relevance. *Plant Ecology & Diversity*, **1**, 273–283.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular ecology*, **16**, 3955–3969.

Jump AS, Peñuelas J (2005) Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters*, **8**, 1010–1020.

Kasprzewska A (2003) Plant chitinases-regulation and function. *Cellular and Molecular Biology Letters*, **8**, 809–824.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Körner C (2003) *Alpine Plant Life*. Heidelberg: Springer, Germany.

Latch EK, Boarman WI, Walde A, Fleischer RC (2011) Fine-scale analysis reveals cryptic landscape genetic structure in desert tortoises. *PloS one*, **6**, e27794.

Manel S, Gugerli F, Thuiller W *et al.* (2012) Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular ecology*, **21**, 3729–3738.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular ecology*, **19**, 3824–3835.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

McCune B, Keon D (2002) Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science*, **13**, 603–606.

Mencuccini M, Piussi P, Zanzi Sulli A (1995) Thirty years of seed production in a subalpine Norway spruce forest: Patterns of temporal and spatial variation. *Forest Ecology and Management*, **76**, 109–125.

Mosca E, Eckert AJ, Di Pierro EA *et al.* (2012) The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular ecology*, **21**, 5530–5545.

Müller-Starck (1995) Genetic variation in high elevated populations of Norway spruce (*Picea abies* (L.) Karst.) in Switzerland. *Silvae Genetica*, **44**, 356–362.

Narum SR, Campbell NR, Kozfkay CC, Meyer KA (2010) Adaptation of redband trout in desert and montane environments. *Molecular ecology*, **19**, 4622–4637.

Narum SR, Hess JE (2011) Comparison of F(ST) outlier tests for SNP loci under selection. *Molecular ecology resources*, **11 Suppl 1**, 184–194.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Neteler M, Bowman MH, Landa M, Metz M (2012) GRASS GIS: A multi-purpose open source GIS. *Environmental Modelling & Software*, **31**, 124–130.

Quantum GIS Development Team (2011) *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project.

Rehfeldt GE (1989) Ecological adaptations in Douglas-Fir (*Pseudotsuga menziesii* var. *glauca*): a Synthesis. *Forest Ecology and Management*, **28**, 203–215.

Ronceret A, Gadea-Vacas J, Guilleminot J, Devic M (2008) The alpha-N-acetyl-glucosaminidase gene is transcriptionally activated in male and female gametes prior to fertilization and is essential for seed development in *Arabidopsis*. *Journal of Experimental Botany*, **59**, 3649–3659.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Sarvas RO (1968) *Investigations on the flowering and seed crop of Picea Abies*. Commun. Inst. For. Fenn., Helsinki.

Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. *Current opinion in plant biology*, **10**, 162–167.

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 595–619.

Saxe H, Cannell MGR, Johnsen Ø, Ryan MG, Vourlitis G (2001) Tree and forest functioning in response to global warming. *New Phytologist*, **149**, 369–399.

Schmidt-Vogt H (1974) Das natürliche Verbreitungsgebiet der Fichte (Picea abies [L.] Karst) in Eurasien. *Allgemeine Forst- und Jagdzeitung*, 145:185–197.

Scotti I, Vendramin GG, Matteotti LS *et al.* (2000) Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified region (SCAR) markers. *Molecular ecology*, **9**, 699–708.

Shen R, Fan J-B, Campbell D *et al.* (2005) High-throughput SNP genotyping on universal bead arrays. *Mutation research*, **573**, 70–82.

Slatkin M, Voelm L (1991) FST in a hierarchical island model. *Genetics*, **127**, 627–629.

Theurillat JP, Guisan A (2001) Potential impact of climate change on vegetation in the European Alps: a review. *Climatic change*, **50**, 77–109.

Thioulouse J, Chessel D, Champely S (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics*, **2**, 1–14.

Tice KA, Carlon DB (2011) Can AFLP genome scans detect small islands of differentiation? The case of shell sculpture variation in the periwinkle *Echinolittorina hawaiiensis*. *Journal of evolutionary biology*, **24**, 1814–1825.

Unger GM, Konrad H, Geburek T (2011) Does spatial genetic structure increase with altitude? An answer from *Picea abies* in Tyrol, Austria. *Plant Systematics and Evolution*, **292**, 133–141.

Vilas A, Pérez-Figueroa A, Caballero A (2012) A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of evolutionary biology*, **25**, 1364–1376.

Weir BS (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358.

Xing S, Lauri A, Zachgo S (2006) Redox regulation and flower development: a novel function for glutaredoxins. *Plant biology (Stuttgart, Germany)*, **8**, 547–555.

CHAPTER 4

Conclusion

This project presented a candidate gene-based approach to explore adaptive genetic variants, at both geographical and regional scale, in natural populations of *Picea abies*, a non-model coniferous species of great relevance from both an ecological and economical point of view. Its wide distribution across many diverse environments renders it an interesting species for investigating the genetic basis of adaptation and identifying the major ecological determinants of plants distribution and adaptive response.

In order to attempt an inspection of the full spectrum of environmental heterogeneity, which leads to adaptive genetic diversity, landscape features and environmental factors, including climate, were analyzed at different spatial scales. Deciphering the interaction between the environment and the adaptation of organisms may be affected by the different spatial scales at which such a phenomenon is investigated (Wiens 1989; Manel *et al.* 2010). For example, climatic effects may be evident at broad scale, while hidden by temporal and/or spatial effects, introduced by local biological interactions, at finer scales. On the other hand these local processes, as for example local topography or edaphic factors controlling plant distribution at local scales, may disappear at broader scales (Levin 1989; Wiens 1989). Therefore, we tried adopting a multi-scale perspective to analyze different

aspects of the interaction between genetic diversity of conifer forests and their natural environment.

To evaluate adaptive responses to climate, a first investigation was conducted at a geographical scale corresponding to the natural distribution of *P. abies* across the Italian Alps. Temperature and precipitation, the two major ecological determinants of plants' natural distribution and adaptation (Berry & Björkman 1980; Manel *et al.* 2012), were investigated for association to potentially adaptive loci.

This analysis revealed a weak population structure across the Italian Alps, with a low overall genetic differentiation (F_{st}-*multilocus* = 0.012), but still within the standard range for conifer species (Hamrick *et al.* 1992; Müller-Starck *et al.* 1992). Similar findings on the genetic differentiation of natural populations of Norway spruce across the Alpine domain were presented also in previous studies with different genetic markers (ALLOZYME: Müller-Starck 1995; Geburek 1998; SCAR: Scotti *et al.* 2000; SSR: Scotti *et al.* 2006; Meloni *et al.* 2007 and SNPs: Heuertz *et al.* 2006; Scalfi *et al. in preparation*). Population structure analyses detected four possible genetic pools of origin highly admixed to form one unique genetic cluster, which includes the isolated population on the Apennine. The only exception was the Valdieri population (our pop. 19), the southernmost population of the western Alps, which represented a disjoint cluster. These results support previous findings (Scotti *et al.* 2000; Meloni *et al.* 2007), which assessed the uniqueness of this population. Therefore, a possible extension of this work could be to investigate another southern indigenous relict population in Alta Pesio Valley in Loc. *Costa del Pari*, together with populations on the French side of the Maritime Alps, in order to have a comparison and possibly find a common gene pool for both populations.

With regards to the ecological determinants, seasonal precipitation and temperature were shown to affect genetic variation at seven loci (UMN_1604_01-348, CL1343Contig1_05-165, 2_374_01-319, 0_18261_01-105, 0_7471_01-399, 0_16480_02-185, CL304Contig1_01-118). The strongest signal came from the association between allele frequencies and seasonal precipitation. In a recent investigation on four alpine conifer species (Mosca *et al.* 2012) we detected similar association patterns, where precipitation appears as the most involved variable in the SNP-climate relationship in *Larix decidua* and *Pinus cembra*. High susceptibility to water shortage in *P. abies* and *P. cembra* has been assessed (Anfodillo *et al.* 1998; Karlsson *et al.* 1997; Ditmarová *et al.* 2010), and different drought avoidance strategies have been observed in the three species (Anfodillo *et al.* 1998). Water availability is one of the major abiotic stressor for conifers; traits related to water shortage stress have revealed a genetic basis in several conifer species and variation at these traits is adaptive (Zhang & Marshall 1994; Olivas-García *et al.* 2000; Eckert *et al.* 2010). In the actual context of global warming, the increase in surface temperatures may have very important consequences for the hydrological cycle, particularly in areas where water supply is strongly determined by melting snow or ice, such as the Alps, and high rate of water shortages are expected (Braun *et al.* 2000; Barnett *et al.* 2005). Therefore the identification and preservation of adaptive polymorphisms in natural populations could be important tools for new management strategies aimed at mitigating the impact of climatic change on the natural forest resources.

However strong, the association between adaptive factors and potential candidate genes depends also on the strength of interpopulation gene flow relative to selection (Lenormand 2002; Savolainen *et al.* 2007), and disentangling the effects of gene flow and selection is not straightforward (Räsänen & Hendry 2008). When selection is stronger than

gene flow, adaptive genetic differences among populations are maintained by a selection-migration balance. In the present study, significant, albeit small, differentiation along the altitudinal gradient was observed, suggesting an effect of selection on just a few SNPs.

The new approach based on Moran's Eigenvectors Maps (MEM) unveiled 19 *loci of ecological relevance* significantly associated to the environmental variation at regional-scale. The majority of these loci were associated to spatial variables representing un-accounted environmental factors and geographical variation, while elevation influenced allele distribution at five SNP loci (CL1308Contig1_03-181, CL1225Contig1_03-91, 2_6313_01-164, 0_8531_01-363, 0_15075_01-341). It is possible that many selective factors so far hidden could be revealed by the application of this new approach. Further studies about the architecture of the traits involved, which are with the greatest probability multifactorial, are needed to assess the relative strength of selection and gene flow, which would countermand the effects of adaptive variation (Volis & Zhang 2010).

An alternative approach is to look for those loci that display excessive differentiation: F_{ST}-based approaches were used in both investigations to detect outlier loci potentially under selection. However, only four loci, significantly detected as outlier in the regional-scale investigation (0_13680_01-216, 2_9845_01-282, UMN_1604_01-348, CL1225Contig1_03-91) found correspondence with the environmental association analyses results. According to Le Corre & Kremer (2012) in some cases F_{ST}-based methods may fail to identify the genetic causes of adaptation, especially for climate related traits, as loci detected as outliers in forest species may harbour low differentiation levels, close to those measured at neutral markers. Therefore, the information brought by loci associated to climatic variables should be considered as valuable as for outlier loci, either

because they are direct targets of selection or because they are genetically linked to a selected locus. (Bonin *et al.* 2006).

Different scales provided different estimates of potentially adaptive loci and only three loci were detected at both spatial scale investigated: one F_{ST}-outlier (2_5483_02-109) and two loci associated to environmental factors (UMN_1604_01-348 and 0_16480_02-185).

Putative functions of detected loci may provide information on their potential role in the adaptation process or response mechanisms to climatic changes. However, knowledge on the functionality of active genomic regions is still strongly limited in conifer trees, mainly due to the large size of their genome. Therefore, it was possible to obtain an informative putative function only for a restrained number of loci; for example, two loci (locus CL304Contig1 and locus 0_18261, both highlighted in the broad scale investigation) are involved in mechanisms active during photosynthesis, therefore strictly affected by the amount of light, temperature, and availability of water. In particular locus CL304Contig1 showed a similarity to a predicted oxygen-evolving enhancer protein 1, a catalytic site within photosystem II (PSII), which is responsible of water oxidation process (Ferreira *et al.* 2004).

This study presented an example of combined approaches for studying adaptive genetic variation of a non-model species in natural environments at different spatial scales. Characterization of local adaptation with putative candidate genes may provide a starting tool for the identification of the adaptive potential of our forest, and thus a valid information to be applied in forest management and conservation, for example by improving predictions of responses to climate change, or even improve seed sources selection for reforestation (Aitken *et al.* 2008). Further developments of this study could

integrate population genomics analysis with common garden experiments on individual phenotypes in order to address the genetic bases of relevant physiological responses to major environmental stresses.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Anfodillo T, Rento S, Carraro V *et al.* (1998) Tree water relations and climatic variations at the alpine timberline: seasonal changes of sap flux and xylem water potential in *Larix decidua* Miller, *Picea abies* (L.) Karst. and *Pinus cembra* L. *Annals of Forest Science*, **55**, 159–172.

Barnett TP, Adam JC, Lettenmaier DP (2005) Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, **438**, 303–309.

Berry J, Bjorkman O (1980) Photosynthetic Response and Adaptation to Temperature in Higher Plants. *Annual Review of Plant Physiology*, **31**, 491–543.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular biology and evolution*, **23**, 773–783.

Braun LN, Weber M, Schulz M (2000) Consequences of climate change for runoff from Alpine regions. *Annals of Glaciology*, **31**, 19–25.

Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular ecology*, **21**, 1548–1566.

Ditmarová L, Kurjak D, Palmroth S, Kmet J, Strelcová K (2010) Physiological responses of Norway spruce (*Picea abies*) seedlings to drought stress. *Tree physiology*, **30**, 205–213.

Eckert AJ, Bower AD, González-Martínez SC *et al.* (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular ecology*, **19**, 3789–3805.

Ferreira KN, Iverson TM, Maghlaoui K, Barber J, Iwata S (2004) Architecture of the photosynthetic oxygen-evolving center. *Science*, **303**, 1831–1838.

Geburek T (1998) Genetic variation of Norway spruce (*Picea abies* [L.] Karst.) populations in Austria. III. Macrospatial allozyme patterns of high elevation populations. *Forest Genetics*, **6**, 201–211.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Heuertz M, De Paoli E, Källman T *et al.* (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics*, **174**, 2095–2105.

Karlsson PE, Medin EL, Wallin G, Selldén G, Skärby L (1997) Effects of ozone and drought stress on the physiology and growth of two clones of Norway spruce (*Picea abies*). *The New Phytologist*, **136**, 265–275.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.

Levin SA (1989) Challenges in the development of a theory of ecosystem structure and function. In: *Perspectives in Ecologicol Theory (ed. J. Roughgarden, R.M. May & S.A. Levin)*, pp. 242–255. Princeton University Press, Princeton, N.J.

Manel S, Gugerli F, Thuiller W *et al.* (2012) Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular ecology*, **21**, 3729–3738.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular ecology*, **19**, 3824–3835.

Meloni M, Perini D, Binelli G (2007) The distribution of genetic variation in Norway spruce (*Picea abies* Karst.) populations in the western Alps. *Journal of Biogeography*, **34**, 929–938.

Mosca E, Eckert AJ, Di Pierro EA *et al.* (2012) The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular ecology*, **21**, 5530–5545.

Müller-Starck (1995) Genetic variation in high elevated populations of Norway spruce (*Picea abies* (L.) Karst.) in Switzerland. *Silvae Genetica*, **44**, 356–362.

Müller-Starck G, Baradat P, Bergmann F (1992) Genetic variation within European tree species. *New Forests*, **6**, 23–47.

Olivas-García JM, Cregg BM, Hennessey TC (2000) Genotypic variation in carbon isotope discrimination and gas exchange of ponderosa pine seedlings under two levels of water stress. *Canadian Journal of Forest Research*, **30**, 1581–1590.

Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology letters*, **11**, 624–636.

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 595–619.

Scotti I, Paglia G, Magni F, Morgante M (2006) Population genetics of Norway spruce (*Picea abies* Karst.) at regional scale: sensitivity of different microsatellite motif classes in detecting differentiation. *Annals of Forest Science*, **63**, 485–491.

Scotti I, Vendramin GG, Matteotti LS *et al.* (2000) Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified region (SCAR) markers. *Molecular ecology*, **9**, 699–708.

Volis S, Zhang Y-H (2010) Separating Effects of Gene Flow and Natural Selection along an Environmental Gradient. *Evolutionary Biology*, **37**, 187–199.

Wiens JA (1989) Spatial Scaling in Ecology. *Functional Ecology*, **3**, 385.

Zhang J, Marshall JD (1994) Population differences in water-use efficiency of well-watered and water-stressed western larch seedlings. *Canadian Journal of Forest Research*, **24**, 92–99.

APPENDIX A

Table S1. Exact location of the 12 Norway spruce provenances used for SNPs identification in *P. abies*, whithin the Comparative Re-Sequencing project (CRSP http://dendrome.ucdavis.edu/NealeLab/crsp/overview).

tree_identifier	elevation	latitude	longitude	country_name
Paab-1	80	52.8667	23.7833	POLAND
Paab-2	80	57.05	23.1667	LATVIA
Paab-3	720	47.35	25.6667	ROMANIA
Paab-4	170	53.4667	26.7167	BELARUS
Paab-5	500	49	24	UKRAINE
Paab-6	850	51.6667	10.6667	GERMANY
Paab-7	700	60.8333	8.61667	Scandinavia
Paab-8	900	60.8333	8.61667	Scandinavia
Paab-9	500	61.8333	11.1	Scandinavia
Paab-10	900	61.7667	11.3833	Scandinavia
Paab-11	300	63	11.5667	Scandinavia
Paab-12	500	60	11.5667	Scandinavia

II

Table S2. Description of the 285 EST unigenes from which 384 single nucleotide polymorphisms (SNPs) where chosen for genotyping via Illumina's BeadXpress hight-throughput platform. Successful SNPs cover 172 (60.4%) EST unigenes.

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
UMN_CL6 Contig1	1	-	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase, expressed	*Oryza sativa Japonica*	NP_001067313	0	Methionine biosynthetic process
UMN_CL415 Contig1	1	-	Uncharacterized protein LOC100248387	*Vitis vinifera*	XP_003634235	1.00E-54	------
UMN_CL132 Contig1	1	-	Malate dehydrogenase, chloroplastic-like	*Vitis vinifera*	XP_002283619	2.00E-161	Cellular carbohydrate metabolic process
UMN_862	1	-	------	------	------	------	------
UMN_853	3	1	Unkown protein	*Picea sitchensis*	ADE76844	3.00E-19	------
UMN_801	1	-	Predicted protein	*Physcomitrella patens* subsp. *patens*	XP_001770015	1.00E-41	------
UMN_7021	2	-	Predicted protein	*Populus trichocarpa*	XP_002324236	4.00E-129	------
UMN_6924	1	-	------	------	------	------	------
UMN_686	2	1	Replication factor C subunit 1-like	*Vitis vinifera*	XP_002265891	8.00E-44	DNA replication
UMN_6852	1	-	------	------	------	------	------
UMN_5384	1	1	Similar to EST GQ03805.B7P1_G15- Active growth	*Picea glauca*	GO367904		------
UMN_5123	1	-	------	------	------	------	------
UMN_501	2	-	BTB/POZ domain-containing protein NPY2	*Vitis vinifera*	XP_002264153	2.00E-21	Response to light stimulus
UMN_4809	1	-	Predicted protein	*Physcomitrella patens* subsp. *patens*	XP_001778594	5.00E-51	------
UMN_4748	2	1	Peptide transporter	*Ricinus communis*	XP_002533893	3.00E-92	Oligopeptide transport

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
UMN_4646	1	-	------	------	------	------	------
UMN_4361	1	1	GRAS family transcription factor	*Populus trichocarpa*	XP_002336555	2.00E-24	Transcription factors putatively involved in tissue development and other processes
UMN_4258	1	-	------	------	------	------	------
UMN_4230	2	-	------	------	------	------	------
UMN_4091	3	3	F-box family protein	*Populus trichocarpa*	XP_002309467	3.00E-18	Ubiquitin-dependent protein catabolic process
UMN_3847	1	1	------	------	------	------	------
UMN_3525	2	-	------	------	------	------	------
UMN_3521	2	1	------	------	------	------	------
UMN_3417	1	-	------	------	------	------	------
UMN_3408	2	2	Nucleosome remodeling factor, p48 subunit	*Physcomitrella patens*	XP_001754225	1.00E-29	------
UMN_3055	1	1	Protein root hair specific 10	*Arabidopsis thaliana*	NP_177203	1.00E-48	------
UMN_2989	1	-	------	------	------	------	------
UMN_2809	1	-	------	------	------	------	------
UMN_2763	1	1	------	------	------	------	------
UMN_2001	1	-	Pre-mrna-processing-splicing factor	*Ricinus communis*	XP_002517654	3.00E-117	Component of spliceosomes, and found to be essential for the catalytic step II in pre-mrna splicing process
UMN_1908	1	-	Beta-adaptin-like protein A	*Vitis vinifera*	XP_002284239	2.00E-124	Plays a role in protein sorting in the late-Golgi/trans-Golgi network (TGN) and/or endosomes.
UMN_1787	1	-	Pentatricopeptide repeat-containing protein	*Vitis vinifera*	XP_002267472	5.00E-27	------
UMN_1604	1	1	SNF2 family DNA-dependent atpase	*Physcomitrella patens* subsp. *patens*	XP_001773860	8.00E-23	Helicase activity, transcriptional regulation, processing of DNA damage

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
UMN_1483	1	-	------	------	------	------	------
UMN_1178	1	1	------	------	------	------	------
UMN_1023	1	-	F-box/LRR-repeat protein 14	*Vitis vinifera*	XP_002272202	1.00E-84	------
CL905 Contig2	2	-	3-oxoacyl-[acyl-carrier-protein] synthase 3 A, chloroplastic-like	*Vitis vinifera*	XP_003631438	4.00E-93	Catalyzes all the condensation reaction of fatty acid synthesis
CL866 Contig1	2	1	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex-like	*Glycine max*	XP_003555893	2.00E-79	------
CL814 Contig1	1	1	Acyl-coenzyme A oxidase 3, peroxisomal	*Ricinus communis*	XP_002525155	1.00E-120	------
CL71Contig1	1	-	Disease resistance associated protein	*Picea abies*	AAV34188	4.00E-180	Resistant to P. Syringae
CL717 Contig1	1	1	Photosystem II core complex proteins psby, chloroplast precursor	*Medicago truncatula*	XP_003601864	1.00E-20	------
CL697Contig 1	1	1	60S ribosomal protein L19, putative	*Ricinus communis*	XP_002516973	2.00E-94	------
CL635 Contig1	1	1	Atpase family AAA domain-containing protein 1-like	*Glycine max*	XP_003544721	2.00E-20	ATPase activity
CL4796 Contig1	1	-	------	------	------	------	------
CL463 Contig2	1	1	Alpha-expansin 8 precursor, putative	*Ricinus communis*	XP_002520863	1.00E-62	------
CL4578 Contig1	1	1	Acetyltransferase, putative	*Ricinus communis*	XP_002531946	1.00E-37	------
CL454 Contig1	1	-	Tubby-like protein	*Physcomitrella patens* subsp. *Patens*	XP_001768253	0	May mediate the ubiquitination and subsequent proteasomal degradation of target proteins
CL4511 Contig1	1	1	Oligopeptidase, putative	*Ricinus communis*	XP_002515011	6.00E-64	------
CL4481 Contig1	1	-	------	------	------	------	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
CL4342 Contig1	1	-	Threonine dehydratase biosynthetic, chloroplastic-like	*Vitis vinifera*	XP_002264311	1.00E-83	Involved in the biosynthesis of isoleucine, and in hydroxamino acid catabolism.
CL4336 Contig1	2	1	RNA recognition motif-containing protein	*Arabidopsis thaliana*	NP_563946	2.00E-29	Protein essential for the binding of mrna to ribosomes
CL4284 Contig1	1	1	------	------	------	------	------
CL4257 Contig1	1	1	Unknown protein. Similar to *Arabisopsis thaliana* At1g73350	*Picea sitchensis*	ABK22187	4.00E-97	------
CL4023 Contig1	1	1	Protein similar to tryptophan synthase, beta subunit region.	*Picea sitchensis*	ABR17785	1.65E-89	Catalyzes the final two steps in the biosynthesis of tryptophan
CL3862 Contig1	2	2	Mitogen activated protein kinase 13	*Pinus taeda*	ACT63867	5.00E-145	Mediators of responses to osmotic shock, oxidative stress, response to cold and involved in anti-pathogen responses. In addition, it is also involved in morphogenesis.
CL3832 Contig1	1	1	Xyloglucan galactosyltransferase KATAMARI1	*Vitis vinifera*	XP_002267390	6.00E-75	Interacts with actin and is required for the proper endomembrane organization and for the cell elongation
CL3795 Contig1	1	1	Amino acid dehydrogenase family protein	*Arabidopsis thaliana*	NP_187837	1.00E-104	------
CL3783 Contig1	1	-	------	------	------	------	------
CL3771 Contig1	2	2	Ubiquitin carrier protein E, ubiquitin-conjugating enzyme E2 32-like	*Glycine max*	XP_003533168	1.00E-118	------
CL3672 Contig1	1	-	------	------	------	------	

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
CL3672 Contig1	1	-	28S ribosomal protein S9, mitochondrial	*Vitis vinifera*	XP_002274146	6.00E-23	Ribosomal protein S9 is one of the proteins from the small ribosomal subunit
CL3602 Contig1	1	1	Protochlorophyllide reductase B	*Zea mays*	NP_001167680	1.00E-16	------
CL3582 Contig1	1	1	-		-	-	------
CL3539 Contig1	1	-	TOM1-like protein 2	*Vitis vinifera*	XP_002275091	2.00E-99	------
CL3507 Contig1	2	1	Ca2+ antiporter/cation exchanger, cation/calcium exchanger 3	*Arabidopsis thaliana*	NP_566474	3.00E-82	------
CL3495 Contig1	1	1	5'-nucleotidase	*Vitis vinifera*	XP_002272482	9.00E-40	Hydrolysis of a nucleotide into a nucleoside and a phosphate.
CL3444 Contig1	2	2	Serine carboxypeptidase-like 27-like isoform 2	*Glycine max*	XP_003556226	4.00E-74	Probable carboxypeptidase
CL3421 Contig1	2	2	------	------	------	------	------
CL3321 Contig1	3	-	Nucleolar protein 56-like	*Glycine max*	XP_003549846	5.00E-106	Protein implicated in ribosomal biogenesis
CL3271 Contig1	1	1	SGHN hydrolase	*Populus trichocarpa*	XP_002327454	3.00E-42	Lipid degradation
CL3162 Contig1	1	1	Ras-related protein Rab11A	*Glycine max*	XP_003525046	4.00E-108	It is associated with both constitutive and regulated secretory pathways, and may be involved in protein transport
CL3148 Contig1	1	1	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1	*Ricinus communis*	XP_002511283	5.00E-112	Controls the expression of genes associated with innate immunity in the absence of pathogens or elicitors. Involved in programmed cell death (PCD) control.
CL3097 Contig1	2	2	------	------	------	------	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
CL304 Contig1	2	1	Oxygen-evolving enhancer protein 1,chloroplastic-like		XP_003557545	1.00E-128	Photosystem II stabilization
CL3036 Contig1	1	1	------	------	------	------	------
CL263 Contig2	1	1	Nuclear acid binding protein	*Ricinus communis*	XP_002510266	5.00E-87	------
CL2637 Contig1	2	2	Beta-1,4-xylosyltransferase IRX10L-like	*Glycine max*	XP_003523179	0	Involved in the synthesis of the hemicellulose glucuronoxylan, a major component of secondary cell walls.
CL2475 Contig1	2	1	Exocyst complex component 8	*Vitis vinifera*	XP_002273667	5.00E-34	This protein interacts with the actin cytoskeletal remodeling and vesicle transport machinery. The complex is also essential for the biogenesis of epithelial cell surface polarity.
CL2310 Contig1	2	-	Asparaginyl-trna synthetase	*Vitis vinifera*	XP_003633115	5.00E-28	ATP + L-asparagine + trna(Asn) = AMP + diphosphate + L-asparaginyl-trna(Asn).
CL2166 Contig1	1	1	Multiple C2 and transmembrane domain-containing protein 2-like, C2 domain-containing protein	*Glycine max*	XP_003554197	1.00E-84	Transferase activity, transferring glycosyl groups.
CL2121 Contig1	1	1	Glycolipid transfer protein	*Ricinus communis*	XP_002528828	8.00E-104	Cytosolic protein that catalyses the transfer of glycolipids between different intracellular membranes.
CL2117 Contig1	1	1	Receptor-like protein kinase HSL1-like	*Glycine max*	XP_003541722	2.00E-66	------
CL1920 Contig1	1	1	Acidic mammalian chitinase-like	*Vitis vinifera*	XP_002263830	3.00E-61	Break down glycosidic bonds in chitin.
CL1905 Contig1	1	1	LIM2 transcription factor	*Pinus pinaster*	ACA33841	5.00E-101	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
CL1888 Contig1	1	-	ATP-dependent Clp protease proteolytic subunit 4	*Glycine max*	XP_003534318	2.00E-102	Plays a major role in the degradation of misfolded proteins. Essential protein required for chloroplast development and integrity.
CL1852 Contig1	1	1	Hypothetical protein containing the poly-adenilate binding domain.	*Populus trichocarpa*	XP_002307526	4.00E-43	Involved in post-transcriptional gene expression processes
CL180 Contig1	1	-	Galacturonosyltransferase 8-like	*Vitis vinifera*	XP_002273962	4.00E-150	May be involved in pectin and/or xylans biosynthesis in cell walls
CL1760 Contig1	1	1	KH domain-containing protein At4g18375-like	*Glycine max*	XP_003523354	2.00E-39	------
CL1758 Contig1	1	-	NADH dehydrogenase ubiquinone iron-sulfur protein 8		NP_178022	2.00E-108	Involed in mitochondrial electron transport, ubiquinone biosynthetic process,metal ion binding.
CL1694 Contig1	3	2	U5 small nuclear ribonucleoprotein component, 116 kd	*Arabidopsis thaliana*	NP_172112	1.00E-99	Involved in regulation of embryo sac egg cell differentiation, embryo development ending in seed dormancy
CL1692 Contig1	1	1	Histone ubiquitination proteins group	*Populus trichocarpa*	XP_002302510	1.00E-85	Involved in chromosome segregation
CL1688 Contig1	2	2	Beta-galactosidase	*Vitis vinifera*	XP_002263382	3.00E-52	The degradation of galactan carried out by this enzyme may determine structural changes and affect cell wall porosity.
CL1669 Contig1	1	-	Predicted protein	*Populus trichocarpa*	XP_002324535	8.00E-73	------
CL163 Contig2	1	-	40S ribosomal protein s3a-like protein	*Elaeis guineensis*	AFS65104	4.00E-159	Component of the ribosome.
CL1581 Contig1	1	-	------	------	------	------	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
CL1530 Contig1	1	1	Histone H2B - performs essential roles in maintaining structural integrity of the nucleosome,	*Arabidopsis thaliana*	NP_197679	1.00E-46	DNA binding, nucleosome assembly
CL149 Contig3	1	-	L-asparaginase	*Pinus sylvestris*	CAK22360	3.00E-158	Catalyzes the hydrolysis of asparagine to aspartic acid
CL1455 Contig1	2	1	Protein binding protein, putative Leucine-rich repeat-containing protein	*Ricinus communis*	XP_002513932	4.00E-109	Protein-protein interactions and have different functions and cellular locations
CL1432 Contig1	1	-	Structural maintenance of chromosomes domain-containing protein	*Citrus sinensis*	AEV43359	8.00E-71	Required for conversion of interphase chromatin into mitotic-like condense chromosomes. Also involved in chromosome segregation in meiosis.
CL1400 Contig1	1	1	30S ribosomal protein S5 (also bacterial type of domain containg protein)	*Arabidopsis thaliana*	NP_180936	2.00E-41	Ribosome, translation, RNA binding, structural component of ribosome, involved in: response to cadmium ion, response to cold
CL1344 Contig1	1	1	6-phosphofructokinase 2	*Arabidopsis thaliana*	NP_199592	2.00E-108	Catalyzes the phosphorylation of fructose-6-phosphate to fructose-1,6-biphosphate, ATP binding, Protein binding
CL1343 Contig1	1	1	Phosphoenolpyruvate carboxykinase 2	*Arabidopsis thaliana*	NP_680468	2.00E-137	Catalyzes the first committed step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis
CL1312 Contig1	3	-	Probable inositol transporter 2-like isoform 1	*Glycine max*	XP_003533841	7.00E-49	Plasma membrane inositol-proton symporter.
CL1308 Contig1	1	1	Unknown protein	*Picea sitchensis*	ABR16813	1.00E-25	------
CL1238 Contig1	2	1	UDP-glucose:glycoprotein glucosyltransferase	*Arabidopsis thaliana*	NP_177278	1.00E-48	Involved in protein glycosylation

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
CL1225 Contig1	1	1	Proline-rich protein	*Pinus taeda*	AAF75825	2.00E-18	Can become covalently cross-linked into the cell wall during development, and may increase wall strength. This cross-linking can also occur rapidly in response to elicitors. Also they may serve as a scaffold for lignin deposition
CL1224 Contig1	1	1	Alpha-n-acetylglucosaminidase	*Medicago truncatula*	XP_003599416	2.00E-46	------
CL1148 Contig1	2	2	Malate dehydrogenase	*Ricinus communis*	XP_002522037	0	Involved in cellular carbohydrate metabolic process.
CL1136 Contig1	1	-	Aspartic proteinase nepenthesin-1	*Vitis vinifera*	XP_002263964	2.00E-92	------
CL1061 Contig1	1	1	Glutathione-s-transferase theta, gst	*Ricinus communis*	XP_002510614	3.00E-48	Involved in cellular detoxification
CL1052 Contig1	1	1	Nitric-oxide synthase	*Arabidopsis thaliana*	NP_190329	2.00E-49	Gtpase/ nitric-oxide synthas
CL1019 Contig1	1	1	Receptor protein kinase, putative leucine rich repeat transmembrane protein kinase	*Arabidopsis thaliana*	NP_199777	2.00E-82	Protein serine/threonine kinase activity
CL1013 Contig1	1	-	Unknown protein	*Picea sitchensis*	ABR18463	8.00E-135	------
CL1004 Contig1	1	1	T-complex protein 1 subunit epsilon-like	*Glycine max*	XP_003538584	9.00E-180	Involved in protein folding, cellular protein metabolic process
2_9940	1	-	Cholesterol transport protein	*Populus trichocarpa*	XP_002312804	5.00E-161	------
2_9845	2	1	Papain family cysteine protease	*Arabidopsis thaliana*	NP_567010	5.00E-84	Endopeptidase with specific substrate preferences.
2_974	1	-	Hypothetical protein 2_974_01	*Pinus labertiana*	AEX12812	2.00E-31	------
2_9665	1	1	Interferon-induced guanylate-binding protein	*Ricinus communis*	XP_002509420	5.00E-86	------
2_9603	1	1	-		-	-	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
2_9466	1	1	Protease ecfe	*Ricinus communis*	XP_002529666	2.00E-23	Cleaves transmembrane domains of substrate proteins, regulating intramembrane proteolysis (RIP) of diverse signal transduction mechanisms
2_9465	1	-	------	------	------	------	------
2_9455	1	1	Pentatricopeptide repeat-containing protein	*Vitis vinifera*	XP_002266469	1.00E-66	------
2_9328	1	1	Transducin/WD40 domain-containing protein	*Arabidopsis thaliana*	NP_190148	6.00E-49	Involved in ucleotide binding and proteine binding, adaptor/regulatory modules in signal transduction, pre-mrna processing and cytoskeleton assembly
2_9280	3	3	Chromatin remodeling complex subunit	*Populus trichocarpa*	XP_002313800	0.056	Zinc ion binding, DNA binding, protein binding
2_9087	2	1	Unknown protein	*Picea sitchensis*	ADE75970	1.00E-79	------
2_8852	2	2	Galactokinase	*Vitis vinifera*	XP_002279647	2.00E-38	Involved in galactose metabolic process
2_8491	1	1	Acyl-coa thioesterase	*Populus trichocarpa*	XP_002301770	5.00e-68	Involved in the regulation of intracellular levels of acyl-coas, free fatty acids and coash, and also in the metabolic regulation of peroxisome proliferation
2_8265	1	-	Cytidine deaminase-like	*Vitis vinifera*	XP_002282373	2.00E-49	Involved in cytidine deamination
2_7803	1	1	Glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family protein	*Arabidopsis thaliana*	NP_567126	4.00E-58	Hydrolase activity, acting on glycosyl bonds,carbohydrate metabolic process, cell wall comp.
2_7725	1	1	Beta-galactosidase 8	*Arabidopsis thaliana*	NP_001189624	4.00E-85	Involved in carbohydrate metabolic process
2_7532	2	1	------	------	------	------	------
2_7351	1	-	Polygalacturonase QRT3-like	*Glycine max*	XP_003534867	4.00E-80	Required for degrading the pollen mother cell wall during microspore development

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
2_7189	1	-	Subtilisin-like protease-like	*Glycine max*	XP_003520892	2.00E-129	Identical protein binding,negative regulation of catalytic activity
2_7025	1	1	Ubiquitin-protein ligase, PUB17	*Selaginella moellendorffii*		2.00E-21	Protein ubiquitination, ubiquitin-protein ligase activity, ubiquitin ligase complex, protein binding
2_6731	2	1	F-box protein GID2	*Ricinus communis*	XP_002510145	2.00E-17	Involved in seed germination, seed dormancy and protein binding
2_6635	4	3	Haloacid Dehalogenase-like Hydrolases	*Pinus radiata*	XP_002527658	1.00E-22	------
2_6491	1	1	Coiled-coil domain-containing protein	*Ricinus communis (XP_002523868), e=3e-10 & max. identity=32%*	AEW08345	1.00E-32	------
2_6368	2	1	------	------	------	------	------
2_6355	1	-	Protein-tyrosine phosphatase mitochondrial 1-like protein	*Vitis vinifera*	XP_002269655	2.00E-34	Protein phosphatase that may mediate dephosphorylation of mitochondrial proteins
2_6313	1	1	Antagonist of mitotic exit network protein 1	*Ricinus communis*	XP_002532563	2.00E-49	Negative regulator of the mitotic exit network (MEN). Acts in the daughter cell to inhibit the mitotic exit pathway once MEN has executed its function. Required for daughter cell separation and chromosome stability. Involved in copper sensitivity
2_6130	1	-	------	------	------	------	------
2_6061	1	-	L-type lectin-domain containing receptor kinase IV.1	*Glycine max*	XP_003529112	4.00E-37	------
2_6052	1	1	Manganese-dependent ADP-ribose/CDP-alcohol diphosphatase	*Vitis vinifera*	XP_003634217	2.00E-83	Hydrolyzes ADP-ribose, IDP-ribose, CDP-glycerol, CDP-choline and CDP-ethanolamine
2_5724	1	-	------	------	------	------	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
2_5668	3	1	GDSL esterase/lipase At3g26430-like	*Glycine max*	XP_003521871	3.00E-72	Involved in lipid metabolic process, hydrolase activity, acting on ester bonds
2_5636	3	2	Pentatricopeptide repeat-containing protein	*Arabidopsis thaliana*	NP_173449	3.00E-38	Binding, cellular component
2_5572	1	-	Zinc finger protein 3-like	*Vitis vinifera*	XP_002282938	1.00E-44	------
2_5483	1	1	Unknown protein	*Picea sitchensis*	ADE76983	1.00E-39	------
2_5073	1	1	Unknown protein (similar to *Arabisopsis thaliana* At1g03900)	*Picea sitchensis*	ABK25067	9.00E-21	Member of the ATP-binding Cassette (ABC) superfamily of membrane transporters
2_4976	2	2	26S proteasome regulatory subunit N6	*Arabidopsis thaliana*	NP_174210.1	2.00E-91	------
2_4892	1	1	Ccaat-binding transcription factor	*Ricinus communis*	XP_002517180	1.00E-74	Involved in chromatin condensation, maintaing structural integrity of nucleosome
2_4723	4	3	Coatomer gamma subunit	*Ricinus communis*	XP_002509477	2.00E-55	------
2_4594	2	1	Nuclear protein zap	*Ricinus communis*	XP_002511527	1.00E-29	------
2_4586	3	1	Oligosaccharyltransferase complex/magnesium transporter family protein	*Arabidopsis thaliana*	NP_849986	1.00E-57	Unknown function, located in plasma membrane, chloroplast
2_4281	2	1	Subtilisin-like protease	*Arabidopsis thaliana*	NP_565330	2.00E-49	Iidentical protein binding,negative regulation of catalytic activity
2_4196	1	1	GTP binding protein	*Arabidopsis thaliana*	NP_569023	3.00E-17	------
2_3947	2	1	AT hook motif DNA-binding family protein	*Ricinus communis*	XP_002519830	3.00E-04	DNA binding
2_3867	3	3	Profilin	*Ricinus communis*	XP_002515952	2.00E-67	Involved in actin cytoskeleton organisation
2_3851	1	1	------	------	------	------	------
2_374	1	1	------	------	------	------	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
2_3726	1	-	Dnaj homolog subfamily B member 13-like isoform 1	*Glycine max*	XP_003527414	8.00E-92	DNAJ heat shock family protein;involved in protein folding
2_3591	1	1	ZF-HD homeobox protein	*Vitis vinifera*	XP_002285709	1.00E-54	Located in nucleus and involved in DNA binding
2_3307	1	1	Pyruvate dehydrogenase E1 component subunit alpha-like	*Glycine max*	XP_003528767	8.00E-30	Has critical function in root hair formation and root development
2_3113	1	-	Dimethyladenosine transferase	*Arabidopsis thaliana*	NP_182264	3.00E-55	Involved in rrna modification, rrna processing
2_3083	2	-	Pyridoxin biosynthesis protein ER1	*Zea mays*	NP_001147020	1.00E-80	Involved in catalytic activity and metabolic activity
2_2960	2	2	Homeobox protein	*Ricinus communis*	(XP_002511608	6.00E-47	Involved in regulation of transcription, DNA-dependent,sequence-specific DNA binding, DNA binding, transcription regulator activity
2_2937	1	1	ATP/ADP transporter	*Poplus trichocarpa*	XP_002298119	3.00E-07	ATP:ADP antiporter activity, ATP binding (integral to membrane)
2_2240	2	1	------	------	------	------	------
2_1528	2	2	------	------	------	------	------
2_10502	2	-	Peptidyl-prolyl cis-trans isomerase B-like	*Vitis vinifera*	XP_002267887	9.00E-111	Ppiases accelerate the folding of proteins. Regulates cell elongation and orientation.
2_10438	1	1	Amidophosphoribosyltransferase, putative	*Ricinus communis*	XP_002513012	5.00E-117	Protein with transferase activity, transferring glycosyl groups
2_10306	1	1	------	------	------	------	------
2_1023	1	1	Hypothetical protein: Ribosomal protein S2 (RPS2)	*Pinus labertiana*	AEW08145	3.00E-63	Involved in the formation of the translation initiation complex, where it might contact the messenger RNA and several components of the ribosome.

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
2_10216	1	1	Ubiquitin-protein ligase activity	*Populus trichocarpa*	XP_002299175	3.00E-46	Cellular protein modification process, protein transport
2_10212	1	-	Glutathione transferase	*Ricinus communis*	XP_002524149	2.00E-110	------
1_6493	1	1	------	------	------	------	------
1_5675	1	-	Lipoxygenase	*Vitis vinifera*	AAO12866	9.00E-119	May be involved in a number of diverse aspects of plant physiology including growth and development, pest resistance, and senescence or responses to wounding
1_3086	1	1	------	------	------	------	------
0_9922	1	-	Unknown protein	*Picea sitchensis*	ACN39982	8.00E-142	------
0_9749	2	1	Serine/threonine-protein kinase PBS1	*Ricinus communis*	XP_002519381	3.00E-18	Protein serine/threonine kinase activity
0_9457	4	3	Pentatricopeptide repeat-containing protein	*Medicago truncatula*	XP_003616196	1.00E-44	Integral membrane, involved in protein transport
0_9389	2	1	Alba DNA/RNA-binding protein , nucleic acid binding protein	*Arabidopsis thaliana*	NP_564325	2.00E-22	Nucleic acid binding
0_9383	1	1	Ubiquitin carboxyl-terminal hydrolase-like protein	*Arabidopsis thaliana*	NP_178033	3.00E-53	Involved in catabolic process, ubiquitin thiolesterase activity.
0_9284	1	1	Triose phosphate/phosphate translocator, non-green plastid,chloroplast precursor	*Ricinus communis*	XP_002525884	2.00E-37	------
0_9119	1	-	NC domain-containing protein	*Arabidopsis lyrata* subsp. *Lyrata*	XP_002884334	5.00E-67	Located in mitochondria
0_9091	1	-	Zinc finger protein	*Medicago truncatula*	XP_003630350	8.00E-69	------
0_9063	1	1	Ubiquitin 11	*Arabidopsis thaliana*	NP_567286	3.00E-120	Protein binding, intercellular
0_8844	1	1	Hypothetical protein: Glycosyltransferase family A (GT-A)	*Pinus taeda*	AFG60469	3.00E-99	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
0_8730	1	-	------	------	------	------	------
0_8683	1	-	Wall-associated kinase, putative	*Ricinus communis*	XP_002524904	2.00E-99	------
0_8675	1	-	Phenylcoumaran benzylic ether reductase	*Populus trichocarpa*	XP_002310455	9.00E-54	Involved in the response to cadmium ion and in the response to oxidative stress
0_8531	2	2	Multicopper oxidase	*Cucumis melo* subsp. *melo*	ADN34061	2.00E-57	Transferase activity, transferring glycosyl groups
0_8479	2	-	Probable inositol transporter 2	*Vitis vinifera*	XP_002278732	3.00E-76	Plasma membrane inositol-proton symporter
0_8359	1	-	Unknown protein	*Picea sitchensis*	ABR18053.1	2.00E-36	------
0_8187	1	-	PRLI-interacting factor A	*Medicago truncatula*	XP_003592814.1	2.00E-31	------
0_8111	1	-	3-hydroxyisobutyrate dehydrogenase	*Arabidopsis thaliana*	NP_194641	5.00E-38	Phosphogluconate dehydrogenase (decarboxylating) activity,
0_7973	2	1	------	------	------	------	------
0_7921	1	1	Short chain dehydrogenase	*Ricinus communis*	XP_002517191	8.00E-45	Oxidoreductase activity
0_7844	1	-	PREDICTED: protein VERNALIZATION INSENSITIVE 3-like	*Glycine max*	XP_003549399	1.00E-30	------
0_7810	1	-	GDSL esterase/lipase At1g74460-like	*Glycine max*	XP_003523018	3.00E-49	Involved in lipid metabolic process
0_7793	1	-	Multidrug resistance protein 1	*Ricinus communis*	XP_002519759	2.00E-36	------
0_7471	2	1	------	------	------	------	------
0_7320	1	-	------	------	------	------	------
0_7171	2	2	------	------	------	------	------
0_6465	1	-	Unknown protein	*Picea sitchensis*	ABK23202.1	9.00E-32	------
0_6421	1	-	NADH dehydrogenase subunit 2	*Cycas taitungensis*	YP_001661394.1	6.00E-20	------
0_6259	1	-	Unknown protein	*Lotus japonicus*	AFK42953.1	2.00E-28	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
0_5786	1	-	Probable receptor-like protein kinase	*Vitis vinifera*	XP_002284588.1	4.00E-57	Involved in protein amino acid phosphorylation
0_5601	1	-	------	------	------	------	------
0_5583	1	-	Unknown protein	*Medicago truncatula*	AFK42641.1	3.00E-78	------
0_5361	1	-	Hypothetical protein SELMODRAFT_408444	*Selaginella moellendorffii*	XP_002967068	3.00E-28	------
0_5038	1	1	Predicted protein	*Poplus trichocarpa*	XP_002302396	4.00E-18	------
0_489	2	2	------	------	------	------	------
0_4829	1	-	Aldose 1-epimerase family protein	*Oryza sativa Japonica*	ABG65914	1.00E-57	Involved in galactose metabolic process, hexose metabolic process and carbohydrate metabolic process
0_4756	1	-	F-box family protein	*Populus trichocarpa*	XP_002328886.1	1.00E-09	Represented in gene networks broadly regulated by microrna-mediated gene silencing via RNA interference
0_4541	1	-	BTB/POZ domain-containing protein	*Arabidopsis lyrata* subsp. *lyrata*	XP_002890464	2.00E-91	------
0_444	1	-	Embryo sac development arrest 6 protein	*Arabidopsis thaliana*	NP_188986	4.00E-05	------
0_4394	1	-	Unknown protein	*Picea sitchensis*	ABR17449.1	3.00E-54	------
0_382	3	1	GDP-fucose protein-O-fucosyltransferase 1	*Pinus taeda*	CAE30295	7.00E-107	------
0_366	2	1	Heat shock factor protein 7	*Zea mays*	NP_001149902	2.00E-56	Involved in the response to stress
0_350	1	1	------	------	------	------	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
0_3415	2	-	ADP-ribosylation factor gtpase-activating protein AGD2	*Arabidopsis thaliana*	NP_176283.1	6.00E-19	Involved in methylation-dependent chromatin silencing, cell-cell signaling, virus induced gene silencing, determination of bilateral symmetry, vegetative phase change, xylem and phloem pattern formation, meristem maintenance, covalent chromatin modification, gene silencing by RNA chromatin silencing by small RNA and flower morphogenesis.
0_3128	2	1	F-box family protein	*Poplus trichocarpa*	XP_002298618	7.00E-18	------
0_3073	1	1	------	------	------	------	------
0_2643	1	1	------	------	------	------	------
0_2433	2	2	Histidine triad family protein	*Arabidopsis lyrata* subsp. *lyrata*	XP_002868118	8.00E-26	Nucleotide binding, located in peroxisome, nucleotide metabloc process
0_2354	1	1	Armadillo repeat only 1 protein	*Arabidopsis thaliana*	NP_195220	1.00E-47	Involved in pollen tube growth, binding, located in cytoplasm and nucleus
0_1949	1	-	Probable pyridoxal biosynthesis protein PDX1.1-like	*Brachypodium distachyon*	XP_003558759	8.00E-64	Can protect cellular membranes from lipid peroxidation.
0_18847	1	1	Subtilisin-like protease	*Vitis vinifera*	XP_002269786.1	1.00E-137	------
0_18619	1	1	------	------	------	------	------
0_18587	1	-	Unknown protein	*Picea sitchensis*	ABK25986.1	1.00E-47	------
0_18439	1	1	Insulin degrading enzyme	*Solanum lycopersicum*	NP_001233926	1.00E-36	Secreted/periplasmic Zn-dependent peptidases, insulinase-like
0_18350	1	-	Phospholipase A1-Igamma3	*Arabidopsis thaliana*	NP_564590	1.00E-65	Involved in triglyceride lipase activity, lipid metabolic process, located in chloroplast
0_18267	1	-	---------	---------	---------	---------	------
0_18261	1	1	NAD(P)-linked oxiduriductase-like protein	*Arabidopsis thaliana*	NP_563947	3.00E-61	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
0_18011	1	1	Ribonucleoside-diphosphate reductase	*Populus trichocarpa*	XP_002307143	6.00E-72	Catalyzes the formation of 2'-deoxyribonucleoside diphosphate from ribonucleoside diphosphate
0_177	1	1	------	------	------	------	------
0_17587	4	4	------	------	------	------	------
0_17368	1	1	------	------	------	------	------
0_17215	2	2	Magnesium-chelatase subunit H, putative	*Ricinus communis*	XP_002532078	1.00E-107	Magnesium chelatase activity, involved in chlorophyll biosynthetic process, and in biosynthetic process
0_1688	1	-	Probable LRR receptor-like serine/threonine-protein kinase At5g10290-like	*Glycine max*	XP_003531345	2.00E-121	------
0_16607	1	1	Multidrug resistance protein ABC transporter family	*Poplus trichocarpa*	XP_002327533	1.00E-62	Atpase activity, coupled to transmembrane movement of substances.
0_16480	1	1	Uncharacterized protein LOC100277866	*Zea mays*	NP_001144794	1.00E-33	------
0_16400	1	-	Unusual floral organs	*Medicago truncatula*	XP_003638299	5.00E-54	------
0_16068	2	-	Sterol 3-beta-glucosyltransferase-like	*Brachypodium distachyon*	XP_003571635.1	1.00E-22	May be involved in decane metabolism and autophagy.
0_15639	1	-	------	------	------	------	------
0_15361	1	-	CASP-like protein 8	*Picea sitchensis*	A9NMM6	2.00E-54	Involved in cell elongation
0_15075	1	1	Grx1 glutaredoxin	*Nicotiana benthamiana*	AER93282	8.00E-69	Obiquitous glutathion (GSH) dependent oxydoreductases. Exert crucial function in the response to oxydative stresses
0_15036	2	1	RING-H2 finger protein ATL78	*Arabidopsis thaliana*	NP_175349	4.00E-24	RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger, Zn binding
0_14976	1	1	------	------	------	------	------
0_14853	1	-	Hat dimerisation domain-containing protein	*Oryza sativa Japonica*	BAI39897	2.00E-56	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
0_14694	1	1	------	------	------	------	------
0_14591	1	-	Cytochrome b6	*Pinus taeda*	ACP51163.1	9.00E-140	This is a component of the complex III (or cytochrome b-c1 complex), which is part of the mitochondrial respiratory chain.
0_1439	1	1	Pre-mrna-splicing factor 38B	*Medicago truncatula*	XP_003616194	1.00E-05	RNA processing, response to salt stress
0_14316	1	1	Ubiquitin-like protein	*Medicago truncatula*	XP_003619702.1	2.00E-111	Involved in protein degradation, cell signaling, activation of protein kinases, and in signaling
0_13978	3	3	Binding protein	*Arabidopsis lyrata subs.lyrata*	XP_002883209	3.00E-51	------
0_13957	3	2	Receptor-like protein kinase HSL1-like (leucine-rich repeat receptor-like protein kinase)	*Glycine max*	NP_001239710	1.00E-50	Protein with serine/threonine kinase activity
0_13929	1	1	GAGA-binding transcriptional activator	*Medicago truncatula*	XP_003589387	1.00E-45	------
0_13913	1	-	Exocyst complex component	*Medicago truncatula*	XP_003588653	4.00E-21	Involved in exocytosis
0_13766	1	-	Protein phosphatase 2c	*Ricinus communis*	XP_002521194.1	1.00E-77	Key component and repressor of the abscisic acid (ABA) signaling pathway that regulates numerous ABA responses, such as stomatal closure, osmotic water permeability of the plasma membrane (Pos), drought-induced resistance and rhizogenesis, response to glucose, high light stress, seed germination and inhibition of vegetative growth.
0_13680	1	1	Hypotetical protein: Vacuolar protein 14	*Pinus Lambertiana*	AEW07885	9.00E-48	------
0_13673	1	-	Unknown	*Picea sitchensis*	ADE77132.1	4.00E-32	------

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
0_13565	2	-	CI small heat shock protein 2	*Prunus salicina*	ACV93249.1	8.00E-55	Involved in the folding and unfolding of other proteins. Its expressions is increased when cells are exposed to elevated temperatures or other stress
0_13383	1	1	Elongation factor EF-2	*Arabidopsis thaliana*	NP_188938.1	3.00E-102	Function in translation factor activity, nucleic acid binding, GTP binding, gtpase activity
0_13305	1	-	Probable protein phosphatase 2C 25 isoform 1	*Vitis vinifera*	XP_002276936.1	9.00E-95	Protein phosphatase that negatively regulates defense respones. Inactivates MPK4 and MPK6 MAP kinases involved in stress and defense signaling
0_13240	2	-	L-aspartate oxidase 1-like	*Vitis vinifera*	XP_002274361.1	3.00E-79	Catalyzes the oxidation of L-aspartate to iminoaspartate
0_13058	1	1	Polygalacturonase	*Medicago truncatula*	XP_003595343	1.00E-102	Involved in cell wall organisation.
0_12329	1	1	PAB8 binding protein 8	*Arabidopsis lyrata subs.lyrata*	XP_002891531	1.00E-46	Nucleic acid binding, RNA binding
0_12190	1	-	PREDICTED: uncharacterized protein LOC100253287	*Vitis vinifera*	XP_002266405	2.00E-99	------
0_12156	1	-	ADP-ribosylation factor gtpase-activating protein AGD3-like	*Glycine max*	XP_003551803	3.00E-90	Involved in the spatial control of provascular differentiation and in auxin signaling but not in polar auxin transport or in auxin responses.
0_12021	1	-	STRUBBELIG-receptor family 6	*Arabidopsis thaliana*	NP_175777	1.00E-32	Involved in tyrosine kinase signaling pathway, and in protein amino acid phosphorylation.
0_11980	2	2	Set domain protein	*Ricinus communis*	XP_002521994	1.00E-77	Involved in chromatin modification, located in nucleus

(Continued)

EST Contig ID	Attempted SNPs	Successfull SNPs[1]	Putative gene product	Best BLAST hit	BLAST GenBank Accession	e-value	Putative function
0_11832	1	-	Probable flavin-containing monooxygenase 1	*Vitis vinifera*	XP_002277560.1	2.00E-29	Required for the establishment of systemic acquired resistance (SAR). Confers a salicylic acid-dependent (SA) resistance to virulent pathogens
0_11772	1	1	O-sialoglycoprotein endopeptidase	*Ricinus communis*	XP_002533361	1.00E-137	Endopeptidase activity, metalloendopeptidase activity, zinc ion binding
0_11649	1	1	TUB8	*Poplus trichocarpa*	XP_00231340	1.00E-147	Structural constituent of cytoskeleton; involved in TUB protein polymerization.
0_11270	1	-	------	------	------	------	------
0_11214	1	1	------	------	------	------	------
0_11090	1	1	Protein OBERON 4-like	*Vitis vinifera*	XP_002274296	XP_0022 74296	Required for the maintenance and/or establishment of both the shoot and root meristems. Involved in the development of the basal pole and in auxin-mediated root and vascular development in the embryo.
0_1099	1	1	Pantoate-beta-alanine ligase	*Arabidopsis thaliana*	AT5G48840.1	5.00E-51	Pantoate-beta-alanine ligase activity.
0_10910	3	3	------	------	------	------	------
0_10754	1	1	FACT complex subunit, global transcription factor group, partial	*Physcomitrella patens* subsp. *patens*	XP_001783705	1.00E-20	Involved in transcription (DNA replication, recombination, and repair), chromatin structure and dynamics
0_10706	1	-	------	------	------	------	------
0_10631	1	1	HSP70T-2; ATP binding	*Arabidopsis thaliana*	AT2G32120.1	2.00E-98	Involved in protein folding, response to high light intensity, response to hydrogen peroxide, response to heat.
0_10515	1	1	------	------	------	------	------
0_10267	2	2	Myb domain protein 55	*Arabidopsis thaliana*	AT1G09540.1	8.00E-71	DNA binding/transcription factor

[1]The 214 snps that were included in the final data-set analyzed

Table S3. Genetic diversity estimates calculated for each successfully genotyped and polymorphic SNP: minor allele frequencies (*MAF*), observed (*Ho*) and expected (*He*) heterozigosity, Wright's inbreeding coefficient (*F*) and *Fst*. ID from L215 to L245 labels the 31 SNPs that failed to pass the quality control step.

ID	SNP	*MAF*	*Ho*	*He*	*F*	*Fst*
L001	UMN_853_01-38	0.2793	0.4222	0.4031	-0.0473	0.0140
L002	UMN_686_01-73	0.1598	0.2663	0.2689	0.0094	0.0157
L003	UMN_5384_02-83	0.0613	0.1201	0.1153	-0.0420	0.0060
L004	UMN_4748_01-38	0.1321	0.2303	0.2296	-0.0030	0.0227
L005	UMN_4361_01-336	0.0660	0.1295	0.1234	-0.0497	0.0048
L006	UMN_4091_02-458	0.1443	0.2512	0.2472	-0.0162	0.0051
L007	UMN_4091_02-39	0.0230	0.0412	0.0450	0.0853	0.0173
L008	UMN_4091_02-137	0.0327	0.0506	0.0634	0.2023	0.0046
L009	UMN_3847_01-252	0.3847	0.4484	0.4740	0.0540	0.0114
L010	UMN_3521_01-170	0.0385	0.0694	0.0741	0.0634	-0.0007
L011	UMN_3408_01-332	0.1059	0.1998	0.1897	-0.0533	0.0183
L012	UMN_3408_01-224	0.1936	0.3653	0.3126	-0.1683	0.0104
L013	UMN_3055_01-224	0.0902	0.1683	0.1643	-0.0241	0.0192
L014	UMN_2763_01-139	0.3232	0.4504	0.4380	-0.0281	0.0081
L015	UMN_1604_01-348	0.2210	0.3454	0.3448	-0.0017	0.0273
L016	UMN_1178_01-83	0.0552	0.1055	0.1043	-0.0106	0.0189
L017	CL866Contig1_01-360	0.4734	0.4600	0.4992	0.0784	0.0292
L018	CL814Contig1_06-287	0.0898	0.1626	0.1637	0.0065	-0.0018
L019	CL717Contig1_05-95	0.0466	0.0880	0.0889	0.0099	0.0116
L020	CL697Contig1_03-204	0.1285	0.2473	0.2242	-0.1028	0.0034
L022	CL463Contig2_02-200	0.0381	0.0642	0.0735	0.1264	0.0159
L024	CL4511Contig1_02-223	0.4732	0.5057	0.4992	-0.0131	0.0144
L025	CL4336Contig1_01-325	0.0175	0.0350	0.0345	-0.0165	0.0044
L026	CL4284Contig1_01-180	0.3398	0.4214	0.4492	0.0619	0.0033
L027	CL4257Contig1_01-391	0.1603	0.2474	0.2696	0.0825	0.0246
L028	CL4023Contig1_01-114	0.1312	0.2285	0.2283	-0.0008	0.0097
L029	CL3862Contig1_06-76	0.4875	0.5396	0.5003	-0.0784	0.0073
L030	CL3862Contig1_06-366	0.0455	0.0862	0.0870	0.0094	0.0004
L031	CL3832Contig1_05-210	0.0102	0.0153	0.0203	0.2432	-0.0035
L032	CL3795Contig1_01-45	0.4640	0.5381	0.4981	-0.0804	0.0204
L033	CL3771Contig1_04-68	0.1624	0.2630	0.2725	0.0348	0.0273
L034	CL3771Contig1_04-419	0.0789	0.1195	0.1456	0.1794	0.0242
L035	CL3602Contig1_03-219	0.0255	0.0461	0.0497	0.0727	0.0117
L036	CL3582Contig1_03-63	0.4207	0.4637	0.4880	0.0499	0.0202
L037	CL3507Contig1_03-191	0.0823	0.1550	0.1513	-0.0244	0.0148
L038	CL3495Contig1_03-187	0.0254	0.0484	0.0496	0.0240	0.0028
L039	CL3444Contig1_02-89	0.1787	0.3035	0.2939	-0.0328	0.0085
L040	CL3444Contig1_02-494	0.1673	0.2621	0.2790	0.0607	0.0114
L041	CL3421Contig1_03-70	0.1467	0.2570	0.2506	-0.0254	0.0095

(Continued)

ID	SNP	*MAF*	*Ho*	*He*	*F*	*Fst*
L042	CL3421Contig1_03-160	0.0118	0.0211	0.0233	0.0957	0.0100
L043	CL3271Contig1_02-86	0.0781	0.1465	0.1442	-0.0162	0.0017
L044	CL3162Contig1_02-56	0.1689	0.3160	0.2811	-0.1242	0.0014
L045	CL3148Contig1_04-86	0.2966	0.4019	0.4178	0.0379	0.0207
L046	CL3097Contig1_01-192	0.2582	0.3539	0.3835	0.0771	0.0147
L047	CL3097Contig1_01-163	0.0127	0.0206	0.0251	0.1810	0.0010
L048	CL304Contig1_01-118	0.2369	0.3183	0.3620	0.1207	0.0509
L049	CL3036Contig1_01-102	0.4760	0.4860	0.4995	0.0270	0.0043
L050	CL263Contig2_03-54	0.3461	0.4303	0.4532	0.0504	0.0119
L051	CL2637Contig1_04-67	0.1114	0.1412	0.1982	0.2876	0.0059
L052	CL2637Contig1_04-145	0.2088	0.3063	0.3308	0.0742	0.0168
L053	CL2475Contig1_02-262	0.0120	0.0240	0.0238	-0.0109	0.0041
L054	CL2166Contig1_01-182	0.4142	0.5230	0.4859	-0.0762	0.0095
L055	CL2121Contig1_07-112	0.3029	0.3733	0.4229	0.1171	0.0057
L056	CL2117Contig1_03-159	0.1121	0.1806	0.1993	0.0940	0.0116
L057	CL1920Contig1_01-146	0.3815	0.4239	0.4726	0.1031	0.0171
L058	CL1905Contig1_03-178	0.4492	0.4915	0.4954	0.0079	0.0150
L059	CL1852Contig1_01-81	0.4679	0.4806	0.4985	0.0359	-0.0070
L060	CL1760Contig1_01-115	0.3765	0.4697	0.4701	0.0007	0.0076
L061	CL1694Contig1_04-90	0.2936	0.4153	0.4153	0.0001	0.0080
L062	CL1694Contig1_01-235	0.2972	0.4177	0.4183	0.0014	0.0127
L063	CL1692Contig1_05-178	0.3838	0.4915	0.4736	-0.0379	0.0065
L064	CL1688Contig1_01-463	0.1816	0.3632	0.2976	-0.2204	0.0009
L065	CL1688Contig1_01-106	0.3846	0.5123	0.4740	-0.0808	0.0198
L066	CL1530Contig1_04-64	0.1435	0.2312	0.2461	0.0602	0.0256
L067	CL1455Contig1_06-124	0.3317	0.4358	0.4439	0.0182	0.0049
L068	CL1400Contig1_02-108	0.1071	0.2114	0.1916	-0.1032	0.0131
L069	CL1344Contig1_03-164	0.0933	0.1867	0.1694	-0.1016	0.0108
L070	CL1343Contig1_05-165	0.3748	0.5315	0.4693	-0.1324	0.0337
L071	CL1308Contig1_03-181	0.4059	0.4794	0.4829	0.0073	0.0209
L072	CL1238Contig1_01-270	0.2034	0.3123	0.3244	0.0373	0.0224
L073	CL1225Contig1_03-91	0.2052	0.3160	0.3266	0.0325	0.0262
L074	CL1224Contig1_01-546	0.1513	0.2567	0.2572	0.0020	0.0066
L075	CL1148Contig1_08-225	0.0838	0.1650	0.1537	-0.0735	0.0241
L076	CL1148Contig1_08-134	0.2996	0.4201	0.4202	0.0003	0.0036
L077	CL1061Contig1_03-147	0.0852	0.1521	0.1560	0.0250	0.0027
L078	CL1052Contig1_03-116	0.1110	0.1811	0.1976	0.0832	0.0127
L079	CL1019Contig1_01-194	0.2611	0.4178	0.3864	-0.0814	0.0011
L080	CL1004Contig1_08-304	0.1891	0.2909	0.3070	0.0525	0.0071
L081	2_9845_01-282	0.1774	0.2868	0.2922	0.0187	0.0104
L082	2_9665_01-175	0.3148	0.3971	0.4319	0.0806	0.0160
L083	2_9603_01-139	0.0267	0.0455	0.0521	0.1278	0.0055
L084	2_9466_01-179	0.1489	0.2756	0.2538	-0.0859	0.0027

(Continued)

ID	SNP	*MAF*	*Ho*	*He*	*F*	*Fst*
L085	2_9455_01-318	0.1097	0.2022	0.1956	-0.0336	-0.0007
L086	2_9328_01-425	0.3633	0.5094	0.4632	-0.0997	-0.0016
L087	2_9280_01-338	0.2743	0.4394	0.3987	-0.1023	0.0064
L088	2_9280_01-193	0.0220	0.0391	0.0430	0.0922	0.0002
L089	2_9280_01-123	0.2972	0.4201	0.4183	-0.0044	0.0036
L090	2_9087_01-39	0.0176	0.0327	0.0345	0.0535	0.0457
L091	2_8852_01-97	0.0801	0.1335	0.1475	0.0952	-0.0063
L092	2_8852_01-381	0.3073	0.4376	0.4262	-0.0266	0.0035
L093	2_8491_01-122	0.2195	0.3321	0.3430	0.0320	0.0135
L094	2_7803_01-235	0.4745	0.4570	0.4993	0.0848	0.0134
L095	2_7725_01-466	0.4831	0.5714	0.5000	-0.1428	0.0053
L096	2_7532_01-155	0.0297	0.0593	0.0576	-0.0293	0.0147
L097	2_7025_01-169	0.0762	0.1524	0.1410	-0.0810	0.0178
L098	2_6731_01-207	0.0936	0.1701	0.1698	-0.0017	0.0133
L099	2_6635_01-85	0.1393	0.2396	0.2401	0.0022	0.0201
L100	2_6635_01-244	0.0297	0.0594	0.0577	-0.0294	0.0029
L101	2_6635_01-164	0.0098	0.0197	0.0195	-0.0087	-0.0066
L102	2_6491_01-360	0.4782	0.4812	0.4997	0.0369	0.0028
L103	2_6368_01-432	0.2215	0.3240	0.3453	0.0615	0.0116
L104	2_6313_01-164	0.1195	0.2100	0.2108	0.0038	0.0077
L105	2_6052_01-165	0.0516	0.0922	0.0980	0.0594	0.0077
L106	2_5668_01-408	0.0591	0.1156	0.1114	-0.0379	0.0022
L107	2_5636_01-399	0.2342	0.3526	0.3592	0.0183	0.0210
L108	2_5636_01-209	0.2712	0.3898	0.3958	0.0150	0.0209
L109	2_5483_02-109	0.1260	0.1676	0.2206	0.2403	0.0421
L110	2_5073_01-488	0.0224	0.0424	0.0438	0.0336	0.0134
L111	2_4976_01-263	0.1509	0.2591	0.2566	-0.0097	0.0014
L112	2_4976_01-176	0.2139	0.3527	0.3367	-0.0475	0.0159
L113	2_4892_01-39	0.3842	0.4633	0.4738	0.0223	0.0073
L114	2_4723_01-90	0.0781	0.1465	0.1442	-0.0162	0.0107
L115	2_4723_01-374	0.0738	0.1356	0.1370	0.0100	0.0142
L116	2_4723_01-276	0.0801	0.1456	0.1475	0.0130	0.0102
L117	2_4594_01-460	0.1262	0.2233	0.2208	-0.0112	0.0121
L118	2_4586_01-365	0.0109	0.0194	0.0216	0.1024	-0.0048
L119	2_4281_02-253	0.1725	0.3034	0.2859	-0.0612	0.0153
L120	2_4196_01-201	0.0253	0.0480	0.0494	0.0293	0.0035
L121	2_3947_01-298	0.0727	0.1358	0.1350	-0.0053	0.0006
L122	2_3867_02-532	0.3139	0.4315	0.4313	-0.0005	0.0015
L123	2_3867_02-440	0.0759	0.1470	0.1405	-0.0463	0.0004
L124	2_3867_02-163	0.1567	0.3134	0.2647	-0.1843	0.0092
L125	2_3851_01-280	0.2488	0.3617	0.3742	0.0336	0.0042
L126	2_374_01-319	0.4086	0.4806	0.4839	0.0067	0.0246
L127	2_3591_03-192	0.4425	0.4245	0.4941	0.1409	-0.0016

(Continued)

ID	SNP	*MAF*	*Ho*	*He*	*F*	*Fst*
L128	2_3307_01-186	0.1025	0.1784	0.1843	0.0320	0.0161
L129	2_2960_02-82	0.0684	0.1247	0.1276	0.0228	0.0068
L130	2_2960_02-335	0.0793	0.1513	0.1462	-0.0351	0.0179
L131	2_2937_01-127	0.0321	0.0569	0.0622	0.0849	-0.0051
L132	2_2240_01-224	0.1512	0.2512	0.2570	0.0226	0.0113
L133	2_1528_01-321	0.3535	0.4407	0.4576	0.0371	0.0188
L134	2_1528_01-235	0.1536	0.1946	0.2603	0.2525	0.0125
L135	2_10438_01-351	0.1543	0.2517	0.2613	0.0367	0.0135
L136	2_10306_01-74	0.0418	0.0788	0.0802	0.0180	0.0073
L137	2_1023_01-130	0.4579	0.4986	0.4972	-0.0028	0.0186
L138	2_10216_01-307	0.2139	0.3554	0.3367	-0.0554	0.0007
L139	1_6493_01-130	0.1226	0.1861	0.2154	0.1359	-0.0027
L140	1_3086_01-101	0.0732	0.1416	0.1359	-0.0421	0.0020
L141	0_9749_01-337	0.1021	0.2016	0.1836	-0.0981	0.0116
L142	0_9457_01-46	0.4658	0.4947	0.4983	0.0072	0.0022
L143	0_9457_01-421	0.3106	0.3877	0.4289	0.0960	0.0123
L144	0_9457_01-115	0.4933	0.5255	0.5005	-0.0499	-0.0005
L145	0_9389_01-134	0.0776	0.1236	0.1433	0.1372	0.0118
L146	0_9383_01-438	0.0726	0.1235	0.1349	0.0845	0.0138
L147	0_9284_02-490	0.1800	0.2848	0.2956	0.0362	0.0141
L148	0_9063_01-370	0.0248	0.0472	0.0485	0.0258	0.0134
L149	0_8844_01-281	0.0107	0.0214	0.0212	-0.0094	0.0230
L150	0_8531_01-363	0.3040	0.4719	0.4238	-0.1135	0.0144
L151	0_8531_01-157	0.2609	0.3935	0.3861	-0.0190	0.0092
L152	0_7973_01-149	0.4915	0.5122	0.5005	-0.0234	-0.0022
L153	0_7921_01-212	0.0771	0.1351	0.1425	0.0520	0.0021
L154	0_7471_01-399	0.2415	0.3425	0.3668	0.0663	0.0432
L155	0_7171_01-359	0.2033	0.3046	0.3243	0.0607	0.0108
L156	0_7171_01-233	0.1289	0.2070	0.2249	0.0795	0.0322
L157	0_5038_01-143	0.0470	0.0915	0.0896	-0.0208	0.0084
L158	0_489_01-316	0.2004	0.3208	0.3208	0.0000	0.0000
L159	0_489_01-166	0.0335	0.0617	0.0648	0.0478	0.0292
L160	0_382_01-139	0.0630	0.1211	0.1181	-0.0249	-0.0036
L161	0_366_02-380	0.1189	0.1723	0.2098	0.1787	0.0219
L162	0_350_01-276	0.2297	0.3479	0.3543	0.0181	0.0184
L163	0_3128_02-79	0.0127	0.0204	0.0252	0.1907	-0.0016
L164	0_3073_01-97	0.1291	0.2170	0.2251	0.0362	0.0267
L165	0_2643_01-338	0.2079	0.3455	0.3297	-0.0477	0.0234
L166	0_2433_01-36	0.2318	0.3523	0.3566	0.0121	0.0047
L167	0_2433_01-290	0.2562	0.3664	0.3816	0.0400	0.0049
L168	0_2354_01-194	0.0663	0.1325	0.1239	-0.0696	-0.0005
L169	0_18847_01-323	0.1344	0.2639	0.2329	-0.1331	0.0103
L170	0_18619_01-261	0.4074	0.4395	0.4834	0.0909	0.0207

(Continued)

ID	SNP	*MAF*	*Ho*	*He*	*F*	*Fst*
L171	0_18439_02-175	0.1429	0.2446	0.2452	0.0026	0.0227
L172	0_18261_01-105	0.3930	0.4535	0.4778	0.0507	0.0231
L173	0_18011_01-360	0.1128	0.2028	0.2004	-0.0122	0.0074
L174	0_177_01-165	0.4230	0.4683	0.4888	0.0419	0.0099
L175	0_17587_01-42	0.1930	0.3131	0.3118	-0.0041	-0.0052
L176	0_17587_01-392	0.1627	0.2716	0.2728	0.0045	-0.0075
L177	0_17587_01-294	0.0576	0.1055	0.1087	0.0294	0.0034
L178	0_17587_01-165	0.0557	0.1041	0.1053	0.0113	0.0106
L179	0_17368_01-52	0.3992	0.4719	0.4803	0.0176	0.0093
L180	0_17215_01-225	0.4648	0.5006	0.4981	-0.0050	0.0094
L181	0_17215_01-108	0.3002	0.4140	0.4207	0.0158	0.0094
L182	0_16607_01-284	0.1480	0.2730	0.2525	-0.0812	0.0119
L183	0_16480_02-185	0.0691	0.1382	0.1288	-0.0729	0.0553
L184	0_15075_01-341	0.1368	0.2358	0.2365	0.0029	-0.0021
L185	0_15036_01-252	0.1438	0.2802	0.2466	-0.1360	0.0130
L186	0_14976_01-305	0.4817	0.5128	0.4999	-0.0257	0.0087
L187	0_14694_01-270	0.3336	0.4362	0.4453	0.0205	0.0087
L188	0_1439_01-226	0.1205	0.1975	0.2123	0.0698	0.0073
L189	0_14316_01-589	0.1830	0.2396	0.2995	0.2000	0.0188
L190	0_13978_01-300	0.2923	0.4315	0.4144	-0.0414	0.0141
L191	0_13978_01-231	0.0841	0.1513	0.1543	0.0193	-0.0014
L192	0_13978_01-102	0.0994	0.1745	0.1792	0.0262	0.0007
L193	0_13957_02-309	0.0461	0.0921	0.0880	-0.0470	0.0181
L194	0_13957_02-27	0.4834	0.4156	0.5002	0.1691	0.0177
L195	0_13929_02-138	0.0589	0.0918	0.1109	0.1721	0.0228
L196	0_13680_01-216	0.4824	0.4564	0.5000	0.0871	0.0283
L197	0_13383_01-139	0.0196	0.0392	0.0385	-0.0185	0.0080
L198	0_13058_01-551	0.0751	0.1356	0.1390	0.0247	0.0129
L199	0_12329_01-89	0.0527	0.0981	0.0999	0.0184	0.0018
L200	0_11980_01-550	0.2094	0.3172	0.3316	0.0433	0.0036
L201	0_11980_01-165	0.0309	0.0546	0.0601	0.0906	-0.0008
L202	0_11772_01-103	0.4097	0.4461	0.4843	0.0789	0.0161
L203	0_11649_01-183	0.0532	0.1063	0.1008	-0.0548	0.0074
L204	0_11214_01-394	0.0599	0.1076	0.1128	0.0460	0.0286
L205	0_11090_01-251	0.1135	0.1978	0.2014	0.0180	0.0065
L206	0_1099_01-242	0.1452	0.2345	0.2485	0.0564	-0.0019
L207	0_10910_02-43	0.1098	0.1905	0.1958	0.0268	0.0282
L208	0_10910_02-239	0.0377	0.0702	0.0727	0.0337	0.0687
L209	0_10910_02-154	0.0536	0.1020	0.1017	-0.0032	0.0402
L210	0_10754_01-320	0.2027	0.3228	0.3236	0.0024	0.0018
L211	0_10631_01-193	0.0256	0.0487	0.0499	0.0238	0.0372
L212	0_10515_01-158	0.2161	0.3402	0.3392	-0.0029	0.0090
L213	0_10267_01-274	0.3855	0.4800	0.4743	-0.0119	0.0094

(Continued)

ID	SNP	*MAF*	*Ho*	*He*	*F*	*Fst*
L214	0_10267_01-148	0.1087	0.1810	0.1941	0.0672	0.0040
L215	UMN_853_01-117	0.0073	0.0146	0.0145	-0.0061	0.0045
L216	UMN_501_01-285	0.4849	0.9588	0.5002	-0.9167	0.0013
L217	UMN_501_01-181	0.4472	0.8943	0.4950	-0.8066	0.0056
L218	UMN_4748_01-308	0.0085	0.0169	0.0168	-0.0073	-0.0021
L219	UMN_1023_01-267	0.0085	0.0121	0.0168	0.2805	-0.0007
L220	CL905Contig2_01-93	0.0048	0.0097	0.0097	-0.0036	-0.0004
L221	CL4481Contig1_04-209	0.4908	0.9816	0.5004	-0.9615	0.0005
L222	CL3770Contig1_01-167	0.0037	0.0073	0.0073	-0.0024	0.0017
L223	CL1758Contig1_04-288	0.4636	0.9133	0.4981	-0.8337	0.0005
L224	CL1238Contig1_01-438	0.0109	0.0121	0.0216	0.4390	0.0205
L225	2_974_01-40	0.4945	0.9842	0.5005	-0.9662	0.0000
L226	2_9087_01-156	0.4982	0.9867	0.5006	-0.9710	0.0001
L227	2_6368_01-296	0.0024	0.0049	0.0049	-0.0012	-0.0025
L228	2_6061_01-241	0.0054	0.0109	0.0108	-0.0043	0.0034
L229	2_3113_01-314	0.0049	0.0097	0.0097	-0.0037	0.0104
L230	2_3083_01-117	0.0091	0.0182	0.0180	-0.0080	0.0071
L231	2_10212_01-304	0.0012	0.0024	0.0024	0.0000	-0.0014
L232	0_9457_01-358	0.3814	0.7629	0.4725	-0.6147	-0.0014
L233	0_9119_01-334	0.0030	0.0061	0.0060	-0.0018	0.0111
L234	0_8683_01-164	0.0042	0.0085	0.0084	-0.0030	0.0488
L235	0_8479_01-390	0.0049	0.0097	0.0097	-0.0037	0.0000
L236	0_7844_01-171	0.0046	0.0092	0.0091	-0.0033	-0.0046
L237	0_7793_01-181	0.0076	0.0152	0.0151	-0.0064	0.0021
L238	0_5361_01-393	0.4846	0.8699	0.5003	-0.7389	-0.0002
L239	0_444_01-334	0.0058	0.0065	0.0116	0.4419	-0.0073
L240	0_366_02-103	0.0030	0.0061	0.0061	-0.0018	0.0001
L241	0_18350_01-179	0.4330	0.8660	0.4917	-0.7613	0.0171
L242	0_1688_02-314	0.0764	0.0109	0.1413	0.9228	0.0102
L243	0_16400_01-384	0.4745	0.9287	0.4994	-0.8595	0.0029
L244	0_13305_02-110	0.0030	0.0061	0.0061	-0.0018	-0.0003
L245	0_12190_02-208	0.0036	0.0073	0.0072	-0.0024	0.0079
	mean	0.1833	0.2666	0.2528	-0.0105	0.0107
	±SD	±0.1535	±0.2090	±0.1670	±0.2011	±0.0114

Table S4. Genetic differentiation estimates (Fst, θ statistics) between pairs of sampled populations tested by 1,000 permutations. In grey are shown non-significant values (P> 0.05).

FST	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0.002	**0.006**	**0.004**	0.004	-0.001	**0.006**	**0.006**	**0.007**	**0.009**	0.004	**0.007**	**0.008**	**0.009**	**0.006**	**0.006**	**0.007**	**0.010**	**0.056**	**0.015**	0.004	**0.006**	**0.007**	**0.030**
2		**0.005**	**0.003**	0.001	0.003	**0.007**	**0.007**	**0.010**	0.002	**0.008**	**0.005**	**0.004**	**0.008**	0.005	**0.008**	**0.009**	**0.008**	**0.062**	**0.013**	**0.008**	**0.009**	**0.010**	**0.031**
3			**0.008**	**0.010**	**0.008**	**0.011**	**0.014**	**0.009**	**0.005**	**0.007**	**0.006**	**0.005**	**0.009**	**0.004**	**0.009**	**0.009**	**0.015**	**0.061**	**0.007**	**0.009**	**0.011**	**0.009**	**0.033**
4				0.001	**0.005**	**0.011**	**0.010**	**0.013**	**0.009**	**0.012**	**0.008**	**0.010**	**0.011**	**0.009**	**0.011**	**0.012**	**0.015**	**0.062**	**0.015**	**0.008**	**0.012**	**0.009**	**0.032**
5					0.002	**0.005**	**0.007**	**0.015**	**0.007**	**0.012**	**0.005**	**0.009**	**0.009**	**0.005**	**0.009**	**0.007**	**0.011**	**0.062**	**0.013**	**0.009**	**0.013**	**0.007**	**0.033**
6						**0.008**	**0.007**	**0.008**	**0.007**	**0.005**	**0.005**	**0.008**	**0.008**	**0.007**	**0.006**	**0.007**	**0.011**	**0.061**	**0.016**	**0.008**	**0.009**	**0.007**	**0.028**
7							0.000	**0.007**	**0.006**	**0.006**	**0.008**	**0.006**	**0.007**	**0.005**	**0.009**	**0.007**	**0.011**	**0.061**	**0.013**	**0.010**	**0.012**	**0.012**	**0.030**
8								**0.007**	**0.008**	**0.008**	**0.006**	**0.007**	**0.008**	**0.008**	**0.006**	**0.005**	**0.012**	**0.064**	**0.021**	**0.017**	**0.019**	**0.019**	**0.031**
9									**0.010**	**0.008**	**0.008**	**0.003**	0.003	**0.007**	**0.007**	**0.007**	**0.009**	**0.059**	**0.016**	**0.016**	**0.015**	**0.015**	**0.030**
10										**0.000**	-0.001	0.006	**0.003**	0.001	**0.005**	0.001	**0.007**	**0.061**	**0.006**	**0.014**	**0.014**	**0.015**	**0.024**
11											0.002	**0.006**	**0.009**	0.000	**0.005**	0.001	**0.011**	**0.068**	**0.013**	**0.011**	**0.016**	**0.010**	**0.034**
12												**0.005**	**0.006**	0.002	**0.004**	0.003	**0.011**	**0.069**	**0.012**	**0.011**	**0.010**	**0.012**	**0.026**
13													**0.002**	**0.004**	**0.006**	**0.004**	**0.010**	**0.052**	**0.011**	**0.007**	**0.009**	**0.008**	**0.030**
14														**0.005**	**0.004**	**0.008**	**0.012**	**0.054**	**0.012**	**0.011**	**0.011**	**0.013**	**0.029**
15															**0.005**	0.001	**0.010**	**0.062**	**0.007**	**0.006**	**0.011**	**0.009**	**0.029**
16																**0.006**	**0.016**	**0.059**	**0.014**	**0.011**	**0.012**	**0.013**	**0.026**
17																	**0.011**	**0.055**	**0.008**	**0.008**	**0.013**	**0.011**	**0.033**
18																		**0.052**	**0.021**	**0.015**	**0.022**	**0.018**	**0.029**
19																			**0.057**	**0.049**	**0.059**	**0.047**	**0.075**
20																				**0.015**	**0.022**	**0.016**	**0.034**
21																					-0.003	0.000	**0.026**
22																						**0.009**	**0.034**
23																							**0.036**
24																							

Table S5. Climatic variables for the 24 sampled populations.

Pop	MAT	AP	maxWmQT	minCQT	WtQP	DQP	MWtQT	MDQT	WmQP	CQP	GDD5
1	6.1	967	29.7	-15.7	379	127	15.1	-2.3	357	170	2023.4
2	4.8	967	27.9	-17.0	379	127	13.8	-3.5	357	170	1733.4
3	3.4	967	25.9	-18.5	379	127	12.4	-4.9	357	161	1444.5
4	5.6	967	28.9	-16.7	379	127	14.7	-2.9	352	164	1932.6
5	4.9	967	28.2	-17.4	379	127	14.1	-3.6	357	170	1817.8
6	4.0	967	27.5	-18.5	379	127	13.2	-4.6	357	171	1668.8
7	8.9	549	30.8	-11.6	225	48	14.5	2.2	196	74	2523.6
8	7.9	549	29.4	-12.4	225	48	13.5	1.1	181	82	2252.9
9	5.4	549	25.6	-14.8	225	48	11.0	-1.2	181	82	1692.5
10	7.7	601	27.4	-11.6	238	50	11.9	1.1	179	80	2076.7
11	4.4	601	26.4	-16.5	238	50	9.3	-3.0	192	95	1641.9
12	6.1	601	25.5	-13.4	238	50	10.4	-0.6	179	95	1750.0
13	2.4	669	26.3	-19.5	247	78	7.8	-4.8	203	125	1419.3
14	4.5	516	24.6	-15.8	208	43	10.2	-1.5	163	74	1563.5
15	2.5	926	25.8	-19.4	339	100	7.1	-5.0	298	158	1434.8
16	3.4	926	24.7	-17.6	339	100	9.6	-3.8	298	158	1436.9
17	4.0	607	24.8	-16.3	254	43	8.3	-3.0	211	78	1437.4
18	6.3	634	26.3	-13.7	241	66	10.2	0.8	180	110	1856.7
19	5.5	1005	29.5	-15.9	441	84	-1.9	11.3	120	285	1931.7
20	0.4	1030	29.8	-23.3	357	154	7.2	-6.5	301	218	1373.9
21	5.5	1775	24.7	-14.3	695	182	7.1	0.7	443	315	1648.4
22	5.8	1473	25.4	-14.5	555	146	7.6	-1.1	442	228	1737.7
23	8.2	1490	27.4	-12.1	584	145	9.6	1.2	391	253	2221.8
24	7.2	1314	25.4	-10.6	616	112	2.0	13.8	157	351	1916.1

*MAT, mean annual temperature (°C); AP, annual precipitation (mm); maxWmQT, maximum warmest quarter temperature (°C); minCQT, minimum coldest quarter temperature(°C); WtQP, wettest quarter precipitation (mm), DQP, driest quarter precipitation (mm); MWtQT, mean wettest quarter temperature (°C); MDQT, mean driest quarter temperature (°C); WmQP, warmest quarter precipitation (mm); CQP, coldest quarter precipitation (mm); GDD5, growing degree days base 5°C.

Table S6. Comparison between average annual precipitations, estimated at each sampling site, across the last 30 years (Avg_AP_30y) and the last 10 years (Avg_AP_10y). Kolmogorov-Smirnov test values (D) and significance (p-value).

Population	Avg_AP_30y	Avg_AP_10y	D	p-value
1	1005	967	0.167	0.985
2	1005	967	0.167	0.985
3	1005	967	0.167	0.985
4	1005	967	0.167	0.985
5	1005	967	0.167	0.985
6	1005	967	0.167	0.985
7	687	549	0.467	0.076
8	687	549	0.467	0.076
9	687	549	0.467	0.076
10	742	601	0.400	0.181
11	742	601	0.400	0.181
12	742	601	0.400	0.181
13	818	669	0.400	0.181
14	706	516	0.400	0.181
15	983	926	0.333	0.375
16	983	926	0.333	0.375
17	659	607	0.300	0.510
18	779	634	0.333	0.375
19	1078	1005	0.300	0.510
20	1033	1030	0.133	0.999
21	1819	1775	0.200	0.925
22	1472	1473	0.133	0.999
23	1537	1490	0.200	0.925
24	1544	1314	0.333	0.375

Table S7. Spearman's rank order correlation coefficient (r_s) values calculated between climatic variable pairs.

	MAT	AP	maxWmQT	minCQT	WtQP	DQP	MWtQT	MDQT	WmQP	CQP	GDD5
MAT	-	-	-	-	-	-	-		-	-	-
AP	-0.05	-	-	-	-	-	-		-	-	-
maxWmQT	0.23	0.06	-	-	-	-	-		-	-	-
minCQT	0.93***	-0.14	-0.06	-	-	-	-		-	-	-
WtQP	0.01	0.97***	0	-0.08	-	-	-		-	-	-
DQP	-0.13	0.91***	0.16	-0.28	0.84***	-	-		-	-	-
MWtQT	0.29	-0.35	0.46*	0.07	-0.29	-0.11	-	-	-	-	-
MDQT	0.89***	-0.03	0.04	0.93***	0.03	-0.25	-0.06	-	-	-	-
WmQP	-0.19	0.6**	0.07	-0.34	0.57**	0.77***	0.19	-0.42*	-	-	-
CQP	-0.03	0.96***	0.1	-0.12	0.94***	0.87***	-0.36	0.02	0.47*	-	-
GDD5	0.94***	-0.03	0.48*	0.76***	0.05	-0.08	0.46*	0.79***	-0.14	0.02	-

*** $p \leq 0.001$; ** $p \leq 0.01$; * $p \leq 0.05$.

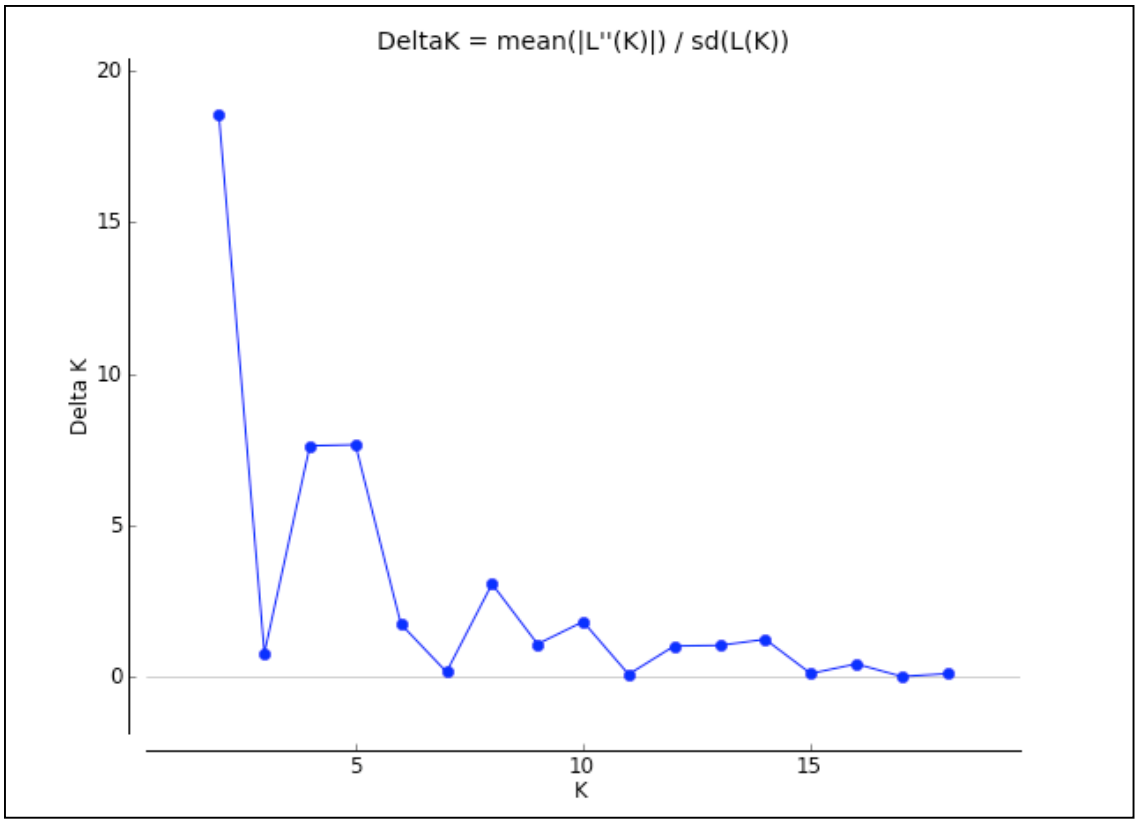

Fig 1S. STRUCTURE v.2.3 results: Evanno's Δ*k*.

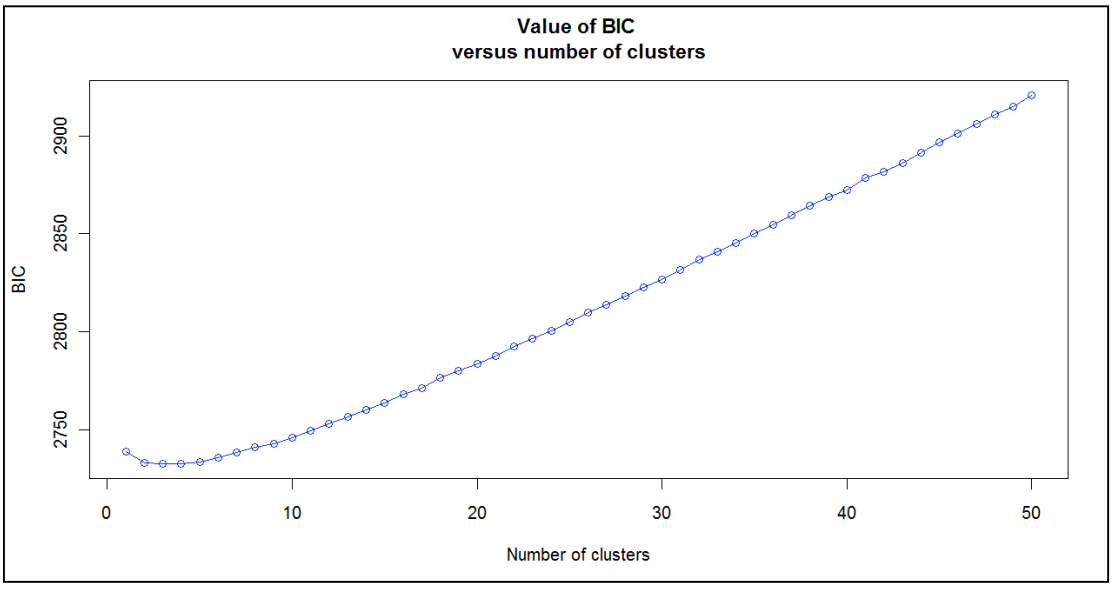

Fig. 2S Estimated BIC at each *k* value inferred by the *k-means* algorithm. Lowest BIC values were reached for *k* = 3 (2732.568) and *k* = 4 (2732.491).

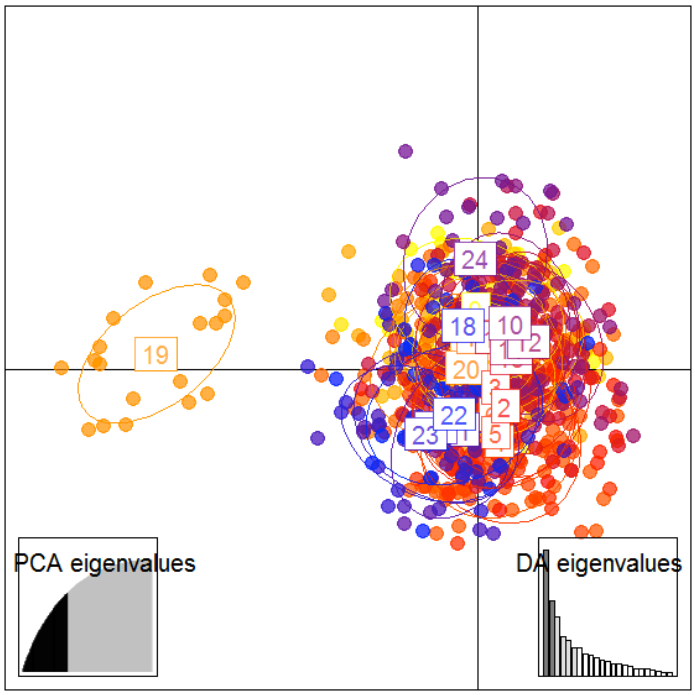

Fig. 3S. DAPC analysis conducted on the entire dataset given the individual populations as prior groups. The proportion of conserved variance was given by 75 PCs retained and is about 70%. Clustering of population is plotted along the first and second axes.

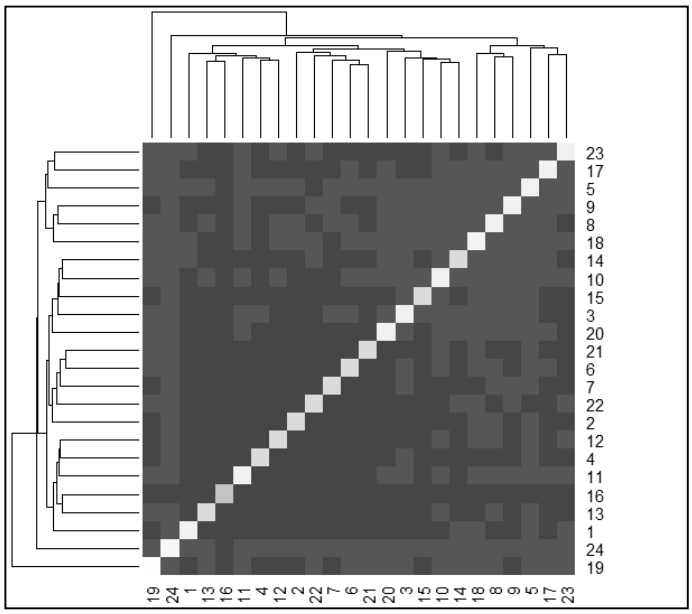

Fig. 4S. Single draw from the posterior of the covariance matrix estimated. Population 19 and 24 are shown to be the most divergent in agreement with population structure analysis results.

APPENDIX B

Table S1. Environmental descriptors and localities for each sampled population

ID	Locality*	n†	Sub-Region	Long	Lat	Altitude (m)	Slope (degree)	Aspect (degree)	Folded-Aspect (degree)
1	Val_Ridanna-l	25	N	11.360	46.887	1109	45.41	32.42	32.42
2	Val_Ridanna-m	25	N	11.348	46.883	1499	16.97	34.99	34.99
3	Val_Ridanna-h	25	N	11.337	46.879	1692	19.22	14.53	14.53
4	Val_Ridanna-l	65	N	11.332	46.903	1242	19.39	276.12	83.88
5	Val_Ridanna-m	25	N	11.336	46.906	1489	38.75	142.59	142.59
6	Val_Ridanna-h	65	N	11.339	46.909	1701	29.31	251.16	108.84
7	Val_Sole-l	25	W	10.938	46.354	775	16.21	334.54	25.46
8	Val_Sole-m	25	W	10.931	46.344	1084	24.28	317.25	42.75
9	Val_Sole-h	25	W	10.940	46.330	1517	25.32	12.2	12.2
10	Val_Calamento-l	35	E	11.488	46.125	1184	32.97	207.55	152.45
11	Val_Calamento-m	25	E	11.491	46.127	1306	35.47	90	90
12	Val_Calamento-h	35	E	11.492	46.132	1455	37.36	114.68	114.68
13	Val_Genova	65	W	10.591	46.195	1683	8.68	34.99	34.99
14	Val_Algone	65	W	10.809	46.146	1757	35.94	133.6	133.6
15	Val_SanNicolo'	65	E	11.766	46.418	1879	29.74	10.08	10.08
16	Paneveggio	65	E	11.764	46.291	1805	12.14	324.46	35.54
17	Passo_Lavaze'	25	E	11.492	46.359	1799	7.26	11.31	11.31
18	Avio	25	W	10.874	45.761	1544	15.69	69.15	69.15

*Elevation plots: l= low, m=medium, h= high; †n: number of individuals sampled in each population

Table S2. List of successfully genotyped SNPs code and SNPs type are provided. NS: non-synonymous SNPs located in coding regions and causing amino acid change; NC: SNPs located in non-coding regions; NA: no information available.

SNP ID	SNP name	SNP type
1	UMN_853_01-38	NA
2	UMN_686_01-73	NA
3	UMN_5384_02-83	NC
4	UMN_4091_02-458	NC
5	UMN_4091_02-39	NC
6	UMN_4091_02-137	NC
7	UMN_3847_01-252	NA
8	UMN_3521_01-170	NA
9	UMN_3055_01-224	NC
10	UMN_2763_01-139	NA
11	UMN_1604_01-348	NS
12	UMN_1178_01-83	NA
13	CL866Contig1_01-360	NC
14	CL814Contig1_06-287	NC
15	CL717Contig1_05-95	NA
16	CL697Contig1_03-204	NC
17	CL635Contig1_01-174	NA
18	CL4578Contig1_02-154	NA
19	CL4511Contig1_02-223	NC
20	CL4336Contig1_01-325	NA
21	CL4284Contig1_01-180	NC
22	CL4257Contig1_01-391	NC
23	CL4023Contig1_01-114	NA
24	CL3862Contig1_06-76	NA
25	CL3862Contig1_06-366	NA
26	CL3832Contig1_05-210	NA
27	CL3795Contig1_01-45	NS
28	CL3771Contig1_04-68	NS
29	CL3602Contig1_03-219	NC
30	CL3582Contig1_03-63	NC
31	CL3495Contig1_03-187	NC
32	CL3444Contig1_02-89	NA
33	CL3444Contig1_02-494	NA
34	CL3421Contig1_03-70	NC
35	CL3421Contig1_03-160	NC
36	CL3271Contig1_02-86	NA
37	CL3162Contig1_02-56	NS
38	CL3148Contig1_04-86	NS
39	CL3097Contig1_01-192	NA
40	CL3097Contig1_01-163	NA
41	CL304Contig1_01-118	NS
42	CL3036Contig1_01-102	NA

(Continued)

SNP ID	SNP name	SNP type
43	CL2637Contig1_04-67	NS
44	CL2637Contig1_04-145	NS
45	CL2475Contig1_02-262	NA
46	CL2121Contig1_07-112	NA
47	CL1920Contig1_01-146	NS
48	CL1905Contig1_03-178	NC
49	CL1852Contig1_01-81	NA
50	CL1760Contig1_01-115	NA
51	CL1694Contig1_04-90	NC
52	CL1694Contig1_01-235	NA
53	CL1692Contig1_05-178	NC
54	CL1688Contig1_01-463	NA
55	CL1688Contig1_01-106	NA
56	CL1455Contig1_06-124	NC
57	CL1343Contig1_05-165	NS
58	CL1308Contig1_03-181	NA
59	CL1238Contig1_01-270	NA
60	CL1225Contig1_03-91	NA
61	CL1224Contig1_01-546	NA
62	CL1148Contig1_08-134	NC
63	CL1019Contig1_01-194	NS
64	2_9845_01-282	NA
65	2_9665_01-175	NA
66	2_9603_01-139	NA
67	2_9466_01-179	NC
68	2_9455_01-318	NS
69	2_9328_01-425	NA
70	2_9280_01-338	NA
71	2_9280_01-193	NA
72	2_9280_01-123	NA
73	2_9087_01-39	NC
74	2_8852_01-381	NS
75	2_8491_01-122	NA
76	2_7725_01-466	NC
77	2_7532_01-155	NA
78	2_7025_01-169	NA
79	2_6635_01-85	NC
80	2_6635_01-244	NC
81	2_6635_01-164	NC
82	2_6491_01-360	NA
83	2_6368_01-432	NA
84	2_6313_01-164	NA
85	2_6052_01-165	NA
86	2_5668_01-408	NA
87	2_5636_01-209	NS
88	2_5483_02-109	NA

(Continued)

SNP ID	SNP name	SNP type
89	2_5073_01-488	NA
90	2_4723_01-90	NA
91	2_4723_01-374	NA
92	2_4723_01-276	NA
93	2_4594_01-460	NC
94	2_4586_01-365	NS
95	2_4281_02-253	NS
96	2_4196_01-201	NS
97	2_3947_01-298	NA
98	2_3867_02-532	NC
99	2_3867_02-440	NC
100	2_3867_02-163	NC
101	2_3851_01-280	NA
102	2_374_01-319	NA
103	2_3591_03-192	NA
104	2_3307_01-186	NA
105	2_2960_02-82	NA
106	2_2960_02-335	NA
107	2_2240_01-224	NA
108	2_1528_01-321	NA
109	2_1528_01-235	NA
110	2_10438_01-351	NA
111	2_10306_01-74	NC
112	2_1023_01-130	NS
113	2_10216_01-307	NA
114	1_6493_01-130	NA
115	1_3086_01-101	NC
116	0_9457_01-46	NA
117	0_9457_01-421	NA
118	0_9457_01-115	NA
119	0_9389_01-134	NC
120	0_9284_02-490	NA
121	0_9063_01-370	NA
122	0_8844_01-281	NC
123	0_8531_01-363	NC
124	0_7973_01-149	NC
125	0_7921_01-212	NA
126	0_7471_01-399	NC
127	0_7171_01-359	NA
128	0_7171_01-233	NA
129	0_5038_01-143	NS
130	0_489_01-316	NA
131	0_489_01-166	NA
132	0_382_01-139	NA
133	0_366_02-380	NA
134	0_350_01-276	NA

(Continued)

SNP ID	SNP name	SNP type
135	0_3128_02-79	NS
136	0_3073_01-97	NS
137	0_2643_01-338	NA
138	0_2433_01-36	NS
139	0_2433_01-290	NC
140	0_2354_01-194	NA
141	0_18847_01-323	NA
142	0_18619_01-261	NC
143	0_18439_02-175	NC
144	0_18011_01-360	NA
145	0_177_01-165	NA
146	0_17587_01-42	NC
147	0_17587_01-392	NA
148	0_17587_01-294	NC
149	0_17587_01-165	NC
150	0_17368_01-52	NC
151	0_16607_01-284	NC
152	0_16480_02-185	NA
153	0_15075_01-341	NS
154	0_15036_01-252	NA
155	0_14976_01-305	NA
156	0_14694_01-270	NC
157	0_1439_01-226	NA
158	0_14316_01-589	NA
159	0_13957_02-309	NA
160	0_13957_02-27	NA
161	0_13929_02-138	NA
162	0_13680_01-216	NC
163	0_13383_01-139	NA
164	0_13058_01-551	NA
165	0_12329_01-89	NA
166	0_11649_01-183	NS
167	0_11214_01-394	NC
168	0_11090_01-251	NS
169	0_1099_01-242	NA
170	0_10910_02-43	NC
171	0_10910_02-239	NC
172	0_10910_02-154	NC
173	0_10515_01-158	NA
174	0_10267_01-274	NS
175	0_10267_01-148	NS

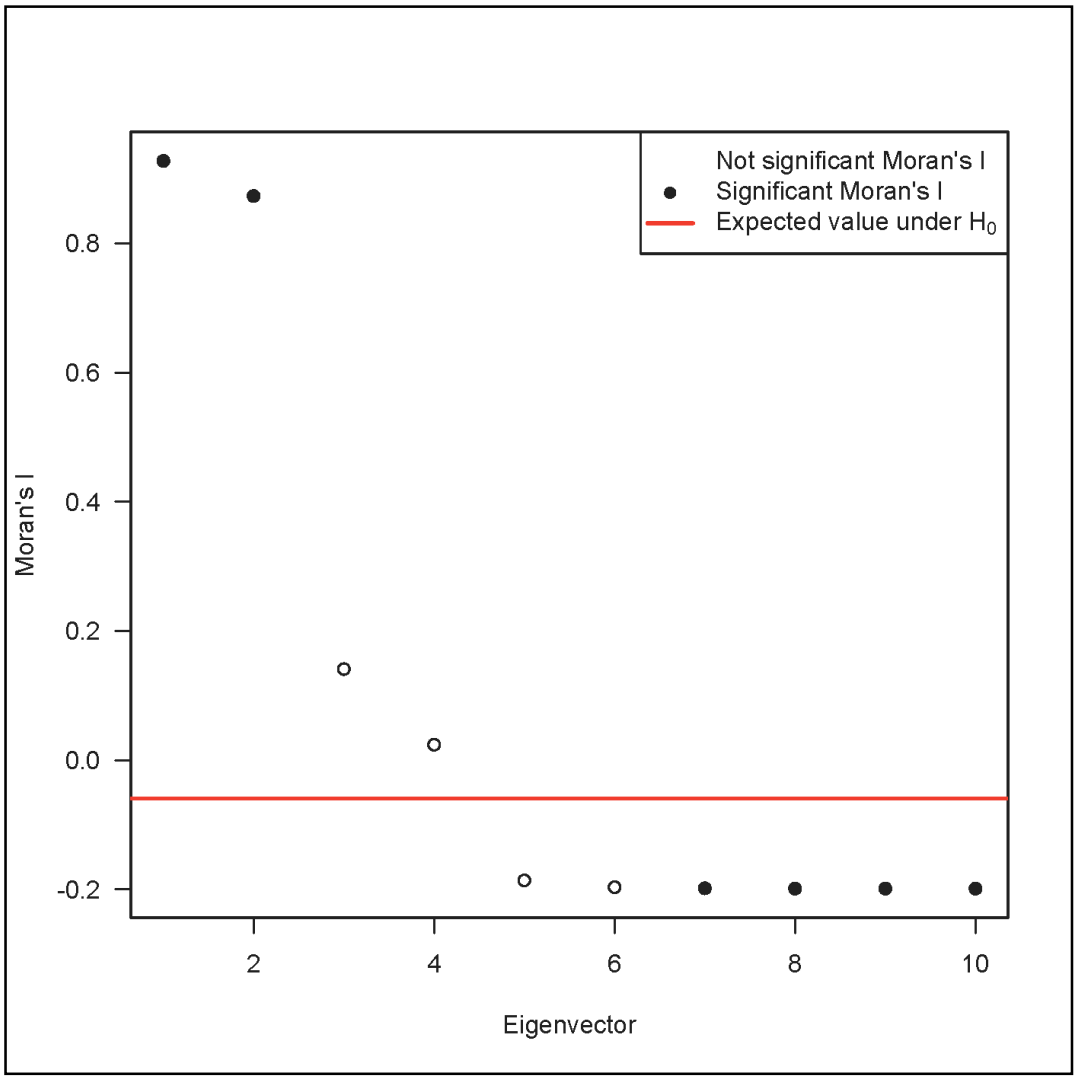

Fig. 1S. Moran's eigenvector map (MEM) variables and their associated Moran's I coefficients of spatial autocorrelation. Only MEMs associated to significant Moran's I were used in the analyses.

Table S3. Twenty-five F_{ST} outlier loci were detected by the Hierarchical Island Model implemented in Arlequin 3.5. Corrected *p-values* by false Discovery Rate (FDR) are reported in table as *q-values*. After FDR correction only locus CL697Contig1_03-204 is still detected as significant outlier. Best informative match for BLAST search in NCBI repository are reported together with the putative gene products.

SNPs name[†]	Ho	F_{ST}	*p-value*	*q-value*	Best BLAST hit	e-value	Putative gene product
CL697Contig1_03-204 [nc]	0.255	0.00001	0.0002	0.0313	*Ricinus communis*	2.00e-94	60S ribosomal protein L19
2_5483_02-109	0.275	0.05823	0.0006	0.0546	*Picea sitchensis*	1.00e-39	Unknown protein
2_9087_01-39 [nc]	0.020	0.00012	0.0033	0.1210	*Picea sitchensis*	1.00e-79	Unknown protein
CL3495Contig1_03-187 [nc]	0.053	0.00019	0.0045	0.1210	*Vitis vinifera*	9.00e-40	Predicted protein: 5' Nucleotidasi
0_7921_01-212	0.135	0.00021	0.0050	0.1210	*Ricinus communis*	8.00e-45	Short chain dehydrogenase
1_6493_01-130	0.205	0.00021	0.0050	0.1210	------	------	------
CL3271Contig1_02-86	0.154	0.00025	0.0060	0.1210	*Populus trichocarpa*	3.00e-42	SGNH_hydrolase
0_13680_01-216 [nc]	0.508	0.04656	0.0062	0.1210	*Pinus lambertiana*	9.00e-48	Hypothetical protein: vacuolar protein14
CL3862Contig1_06-366	0.074	0.00026	0.0062	0.1210	*Pinus taeda*	5.00e-145	Mitogen activated protein kinase 13
0_10267_01-148 [ns]	0.222	0.00031	0.0076	0.1332	*Arabidopsis thaliana*	8.00e-71	Myb domain protein 55
0_350_01-276	0.338	0.00036	0.0088	0.1338	------	------	------
CL1238Contig1_01-270	0.302	0.00037	0.0092	0.1338	*Arabidopsis thaliana*	1.00e-48	UDP-glucose: glucosyltransferase
2_9845_01-282	0.292	0.00056	0.0141	0.1893	*Arabidopsis thaliana*	5.00e-84	Papain family cysteine protease
CL1225Contig1_03-91	0.354	0.00071	0.0173	0.2158	*Pinus taeda*	2.00e-18	Proline-rich protein related to cereal-type Alpha-amylase inhibitors and lipid transfer proteins.
2_5073_01-488	0.031	0.03390	0.0252	0.2600	*Picea sitchensis*	9.00e-21	Unknown protein
0_16480_02-185	0.082	0.00106	0.0255	0.2600	*Zea mays*	1.00e-33	Uncharacterized protein LOC100277866
0_9457_01-421	0.436	0.03686	0.0261	0.2600	*Medicago truncatula*	1.00e-44	Pentatricopeptide repeat-containing protein
UMN_1604_01-348 [ns]	0.360	0.03696	0.0267	0.2600	*Physcomitrella patens subsp.patens*	8.00e-23	SNF2 family DNA-dependent ATPase
2_4281_02-253 [ns]	0.284	0.03455	0.0288	0.2649	*Arabidopsis thaliana*	2.00e-49	Subtilisin-like protease
2_3867_02-440	0.144	0.00153	0.0372	0.3227	*Ricinus communis*	2.00e-67	Profiling protein
CL2475Contig1_02-262	0.034	0.00147	0.0392	0.3227	*Vitis vinifera*	5.00e-34	Exocyst complex component 8
0_18011_01-360 [nc]	0.216	0.00166	0.0406	0.3227	*Populus trichocarpa*	6.00e-72	Ribonucleoside-diphosphate reductase
2_8491_01-122	0.352	0.00182	0.0444	0.3234	*Populus trichocarpa*	5.00e-68	Acyl-CoA thioesterase
0_11090_01-251[ns]	0.224	0.00184	0.0455	0.3234	*Vitis vinifera*	2.00e-34	Protein OBERON 4-like
CL1920Contig1_01-146 [ns]	0.474	0.03230	0.0497	0.3234	*Vitis vinifera*	3.00e-61	Acidic mammalian chitinase-like

[†]Whenever information was available SNP type code was reported: ns: non-synonymous SNPs located in a coding region and causing amino acid change; nc: SNPs located in a non-coding region.

Table S4. Loci identified by multiple linear regressions as putatively under selection. Significant adjusted R^2 and *p-value* are reported (significant *p-value* threshold was set to 0.1, after Holm correction). Best informative match of BLAST search in NCBI repository are reported together with the putative gene products

SNP loci[†]	R^2_{adj}	*p-value*	Best BLAST hit	e-value	Putative gene product
UMN_1604_01-348[ns]	0.56	0.072	*Physcomitrella patens subsp.patens*	8.00e-23	SNF2 family DNA-dependent ATPase
CL4257Contig1_01-391[nc]	0.66	0.030	*Picea sitchensis*	4.00e-97	Unknown protein.
CL4023Contig1_01-114	0.60	0.026	*Picea sitchensis*	1.65e-89	Protein similar to tryptophan synthase, beta subunit reg.
CL1688Contig1_01-106	0.73	0.006	*Vitis vinifera*	3.00e-52	Beta-galactosidase
CL1343Contig1_05-165[ns]	0.65	0.041	*Arabidopsis thaliana*	2.00e-137	Phosphoenolpyruvate carboxykinase 2
CL1308Contig1_03-181	0.82	0.013	*Picea sitchensis*	1.00e-25	Unknown protein
CL1225Contig1_03-91	0.70	0.014	*Pinus taeda*	2.00e-18	Proline-rich protein related to cereal-type alpha-amylase inhibitors and lipid transfer proteins.
CL1224Contig1_01-546	0.53	0.085	*Medicago truncatula*	2.00e-46	Alpha-N-acetylglucosaminidase (NAGLU family)
2_9845_01-282	0.51	0.072	*Picea sitchensis*	5.00e-107	Papain family cysteine protease
2_6635_01-85[nc]	0.66	0.014	*Pinus radiata*	1.00e-32	Hypothetical protein: Haloacid Dehalogenase-like Hydrolases
2_6313_01-164	0.61	0.065	*Ricinus communis*	2.00e-49	Conserved hypothetical protein: Antagonist of mitotic exit network protein 1
2_4586_01-365[ns]	0.60	0.027	*Arabidopsis thaliana*	1.00e-57	Oligosaccharyltransferase complex/magnesium transporter family protein
2_4196_01-201[ns]	0.82	0.011	*Arabidopsis thaliana*	3.00e-17	GTP binding protein
2_1023_01-130[ns]	0.54	0.004	*Pinus lambertiana*	3.00e-63	Hypothetical protein: Ribosomal protein S2 (RPS2)
0_8844_01-281[nc]	0.57	0.085	*Pinus taeda*	3.00e-99	Hypothetical protein: glycosyltransferase family A (GT-A)
0_8531_01-363[nc]	0.67	0.002	*Cucumis melo subsp. melo*	2.00e-57	Multicopper oxidase
0_7973_01-149[nc]	0.49	0.002	------	------	------
0_15075_01-341[ns]	0.56	0.050	*Nicotiana benthamiana*	8.00e-69	GRX1 Glutaredoxin
0_13680_01-216[nc]	0.58	0.017	*Pinus lambertiana*	9.00e-48	Hypotetical protein: Vacuolar protein 14

[†]Whenever information was available SNP type code was reported: ns: non-synonymous SNPs located in coding regions and causing amino acid change; nc: SNPs located in non-coding regions